A molecular genetic analysis of Crohn's disease susceptibility loci in psoriasis

Quaranta, Maria

Awarding institution:
King's College London

A molecular genetic analysis of Crohn's disease susceptibility loci in psoriasis

Thesis submitted for the degree of
Doctor of Philosophy at King's College London

By

Maria Quaranta

King's College London,
School of Medicine,
Division of Genetics and Molecular Medicine
Department of Medical and Molecular Genetics
Guy's Hospital, Tower Wing, 9th Floor,
London, SE1 9RT.

March 2012

Declaration

The work presented here is my own and all experiments were performed by me. The only exceptions are the genotyping of SNPs rs11747270, rs10210302, rs2301436 (carried out by Ms. Natalie Wolf), the establishment and stimulation of primary keratinocyte cultures (performed by Dr. John Mee, Dr. Francesca Capon and Dr. Paola Di Meglio) and the analysis of the Collaborative Association Study of Psoriasis transcriptome data (carried out by Ms. Chrysanthi Ainali).

Maria Quaranta

Abstract

Psoriasis is an immune-mediated skin disorder that is inherited as a complex trait. Genome-wide linkage and association studies have identified a major disease susceptibility locus (*PSORS1*) and several genetic determinants of smaller effect. At least two of these (the *IL12B* and *IL23R* genes) have independently been associated with Crohn's disease (CD). Thus, the aim of this project was to investigate the genetic overlap between psoriasis and CD, with a view to identifying shared pathogenic pathways.

In the first phase of the study, 26 CD variants were genotyped in 1,256 psoriatic patients and 2,938 unaffected individuals. Significant associations (FDR < 0.01) were observed for three markers, mapping to chromosomes 6p22, 21q21 and 21q22. The analysis of an independently ascertained dataset (1,348 cases *vs*. 1,368 controls) validated the chromosome 6p22 association, with the critical SNP (rs6908425) yielding a combined *P* value of 4×10^{-6}. This marker lies within the *CDKAL1* gene, which also harbours type 2 diabetes (T2D) associated alleles.

Since the mechanisms mediating the pathogenic effect of *CDKAL1* are poorly understood, an investigation into gene function was undertaken in the second part of the study. Real-time PCR analyses of multiple cDNA panels showed that *CDKAL1* is abundantly expressed in immune cells, especially in $CD4^+$ and $CD19^+$ lymphocytes. Stable *CDKAL1* knock-down cell lines generated by sh-RNA lentiviral transduction were also analysed. This showed that *CDKAL1* silencing results in reduced cell proliferation and altered cell cycle progression. Transcription profiling of the knock-down cells identified a number of differentially expressed genes, mostly involved in housekeeping functions and inflammatory responses.

Taken together, these findings indicate that *CDKAL1* is a pleiotropic gene conferring

susceptibility to psoriasis, CD and T2D. The results of functional studies suggest that this pathogenic role may be mediated by an effect on inflammatory responses and cell cycle progression.

Acknowledgments

I wish to thank all the people who helped and supported me during my PhD.

I would like to express my sincere gratitude and my thanks to my first supervisor Dr. Francesca Capon for offering me a PhD position, for being patiently by my side during these years, guiding and supporting me through my PhD and also for her continuous challenges that have contributed to my personal and professional development.

I would like to thank my second supervisor Prof. Richard C. Trembath for giving me the opportunity to move my first research steps in his laboratory and for his helpful input into my research. My gratitude also goes to my third supervisor Prof. Jonathan Barker for his support, for sponsoring my attendance to scientific conferences and for helping the small Italian community of the 9th floor to feel at home providing us with an Espresso machine!

Additionally, I would like to thank Natalie Wolf for taking the time to show me the ropes in the lab and Dr. Dimitra Dafou for her supervision during the last months of my PhD. I would also like to show my gratitude to our great lab manager Bethan Johnes for helping me during my last months in London and to Dr. Alex Clop for his friendship, support, and help with bioinformatics. I also would like to thank Dr. Massimo Mangino for his contribution to the statistical analysis of CASP data.

I could not have done this work without the presence of my first mentor and good friend Dr. Paola Di Meglio. Over seven years of friendship she has always been by my side, supporting and encouraging me in all my decisions and sometimes crazy ideas. I am sincerely grateful for the lab training that she gave me during my Master degree and later on, for her friendship and for sharing with me my first experience abroad, with all the happy and difficult moments. Thank you for being there every time I look for you.

I would like to thank my colleague and good friend Anna Bertoni for her sincere friendship, for providing me with all the support and help that I needed, including the strength that she gave me to get over difficult moments, and for the nice evenings in the pub.

I would also like to thank my friend and colleague Luca Napolitano for his help with the immunofluorescence and imaging and for all the nice moments we have spent together.

I would also like to thank my flat mates and friends Mimi Poon, Manuela Terranova Barbeiro, the “Italian Family” aka Dr. Elisa Binda and Dr. Michele Miragoli, Elisabetta Botti, Antonella Di Cesare, Alessandra Marelli, Isabella Tosi and Federica Villanova for believing in me and for making my time in London memorable. I would also like to thank Viviana and Davide, for being my friends during these stressful months.

I thank all my colleagues in the Trembath group and my friends in the Department of Medical and Molecular Genetics & St. John’s Institute of Dermatology for their constant support and for all the fun we have had in these years.

In addition, I would like to acknowledge The Psoriasis Association for funding my PhD research.

I would like to thank Dr. Matt Arno for all his help with the analysis of microarray data, for patiently answering all my questions and for replying to countless e-mails and phone calls. I also wish to thank Dr. Susanne Heck and PJ Chana for their technical assistance with flow cytometry experiments.

Finally, I would also like to express my sincere gratitude to Prof. Carsten Schmidt-Weber, Dr. Stefanie Eyerich and Dr. Kilian Eyerich for giving me the opportunity to start working in their group while I was writing up my thesis and for fully supporting

me in this time.

Grazie a mia madre e a mio padre per l'educazione che mi hanno dato e per essere sempre al mio fianco risollevandomi anche nei momenti più difficili.

Un grazie speciale va a Rita e, soprattutto, a mio fratello Nicola: punto fermo della mia vita, unica persona che è riuscito e riesce a farmi ragionare anche nei miei momenti più irrazionali e di cui non potrei fare a meno.

Questa tesi e questo tempo della mia vita è dedicato a voi.

TABLE OF CONTENTS

Declaration……..…….……I

Abstract……..…….……...….II

Acknowledgements……..…….….IV

List of figures……..…….……...XIV

List of tables……..…….....…XVII

Abbreviations……..…….……...XIX

Chapter 1: General introduction……..……..…….……...1

1.1 The immune system……..……...1

1.1.1 Innate immunity……..…..1

1.1.1.1 Innate immune receptors……1

1.1.1.1.1 Toll-like receptors……...1

1.1.1.1.2 RIG-I like receptors…….…2

1.1.1.1.3 NOD-like receptors…….…2

1.1.1.2 Innate immune cells……...3

1.1.1.2.1 Keratinocytes……...3

1.1.1.2.2 Macrophages and neutrophils……..3

1.1.1.2.3 Natural Killer cells……..4

1.1.1.2.4 Dendritic cells…….5

1.1.2 Adaptive immunity……...7

1.1.2.1 B lymphocytes and the humoral response……..7

1.1.2.2 T lymphocytes and antigen presentation……8

1.1.2.2.1 T helper cell subsets……8

1.2 The structure of skin……..……13

1.2.1 The epidermis……..13

1.2.2 The dermis……...14

1.2.3 Skin resident immune cells…….14

1.3 The gastrointestinal mucosa……..16

1.3.1 Histology of the gastrointestinal tract…….……16

1.3.2 The mucosal barrier……19

1.3.3 The mucosal immune system……..19

1.4 Identification of complex disease susceptibility genes……..20

1.4.1 The genetic component of complex traits……...20

1.4.2 Genome-wide linkage studies…….20

1.4.3 Genome-wide association studies (GWASs)…….21

1.4.3.1 The HapMap Project……21

1.4.3.2 The 1,000 Genome Project……...22

1.4.3.3 Experimental design of GWASs……..22

1.4.3.4 Successes and limitations of GWASs……...23

1.4.4 Follow-up of association signals…….23

1.4.4.1 Fine mapping of disease susceptibility intervals.…….24

1.4.4.2 Functional studies……24

1.5 Psoriasis……..……...26

1.5.1 Clinical presentation……...26

1.5.2 Immunopathogenesis……..29

1.5.3 Genetics of psoriasis……...32

1.5.3.1 The *PSORS1* locus…….…..32

1.5.3.2 Non-MHC disease susceptibility loci……..35

1.6 Crohn's disease……..……40

1.6.1 Clinical presentation……...40

1.6.2 Immunopathogenesis……..43

1.6.3 Genetics of Crohn's disease……45

1.6.3.1 The *NOD2/CARD15* susceptibility locus…….45

1.6.3.2 Other Crohn's disease susceptibility loci…….45

1.7 Aim and overview of the project…….……56

Chapter 2: Materials and methods……..57

2.1 Materials……57

2.1.1 General reagents…….57

2.1.2 General enzymes and related buffers……..60

2.1.3 Electrophoresis gel and loading buffer constituents.……..61

2.1.4 Bacterial and tissue culture growth medium constituents……..62

2.1.5 Cytokines……63

2.1.6 Antibodies……...63

2.1.7 Molecular biology kits……64

2.1.8 Stock solutions……65

2.2 Subjects……..68

2.2.1 Study subjects…….68

2.2.2 DNA extraction from peripheral blood……...68

2.2.3 DNA plating……69

2.3 DNA typing……69

2.3.1 Single Nucleotide Polymorphism genotyping……69

2.3.2 Direct sequencing……...70

2.3.2.1 Primer design……...70

2.3.2.2 Polymerase Chain Reaction (PCR)..……72

2.3.2.3 Direct sequencing…….……72

2.4 Cell culture……73

2.4.1 Culture conditions……...73

2.4.1.1 Adherent immortalized cell lines……....…….73

2.4.1.2 Cell lines grown in suspension…….……74

2.4.1.3 Primary keratinocytes……..74

2.4.2 Immunofluorescence staining…….75

2.5 Cell transfection and transduction…….75

2.5.1 Transient *CDKAL1* RNA interference……75

2.5.2 Stable *CDKAL1* RNA interference…….76

2.5.2.1 Lentivirus production……...76

2.5.2.2 Cell transduction……..76

2.5.2.3 HaCaT cell clonal expansion……...77

2.5.2.4 Jurkat T cell clonal expansion……..77

2.6 Cell cycle analysis…….78

2.6.1 WST-8 proliferation assay……..78

2.6.2 Propidium iodide staining……...78

2.7 Plasmid DNA cloning……79

2.7.1 Preparation of DH5α competent cells…….79

2.7.2 Digestion…….79

2.7.3 DNA ligation and bacterial transformation…….82

2.7.4 Colony screening……82

2.7.5 Plasmid DNA extraction…….……83

2.8 Transcript analysis…….83

2.8.1 RNA extraction……...83

2.8.2 cDNA and cRNA synthesis…….……83

2.8.3 Hybridisation of gene expression chips……..84

2.8.4 Qualitative Reverse Transcription (RT)-PCR…….84

2.8.5 Quantitative RT-PCR……..84

2.9 Protein analysis……..84

2.9.1 Total cell lysis of adherent cells……..84

2.9.2 Total cell lysis of suspension cells……..85

2.9.3 Nuclear and cytoplasmic protein extraction……...85

2.9.4 Protein quantification……..86

2.9.5 Western blotting……..86

2.9.6 Isopropyl-β-D-1-thiogalactopyranoside (IPTG) induction…….89

2.10 Statistical analyses……...89

2.10.1 Power calculations……89

2.10.2 Genotype imputation……90

2.10.3 Analysis of case-control data……90

2.10.4 Analysis of linkage disequilibrium conservation……..90

2.10.5 Multiple testing corrections……..91

2.11 Bioinformatics……91

2.11.1 Evolutionary conservation analysis……..91

2.11.2 Analysis of microarray data……..91

2.11.3 Analysis of real-time PCR data……92

Chapter 3: Genetic analysis of 26 Crohn's disease (CD) susceptibility loci in psoriasis……..96

3.1 Association analysis……..……..96

3.2 Characterization of three psoriasis associated loci……..103

3.3 Replication analysis of disease associated SNPs……108

3.4 Analysis of allelic heterogeneity at the *CDKAL1* locus……...112

3.5 Analysis of type 2 diabetes susceptibility SNPs in the psoriasis dataset…….116

3.6 Discussion……120

Chapter 4: Analysis of *CDKAL1* gene expression…….122

4.1 The *CDKAL1* locus……..122

4.2 Experimental analysis of *CDKAL1* transcript isoforms……..124

4.3 Analysis of *CDKAL1* tissue distribution…….126

4.4 Discussion……132

Chapter 5: Functional analysis of the CDKAL1 protein…….135

5.1 Bioinformatics…….135

5.2 Validation of a α-CDKAL1 antibody……...140

5.3 CDKAL1 subcellular localization……...147

5.4 Generation of stable *CDKAL1* knock-down cell lines…….150

5.5 Cell cycle analysis of *CDKAL1* knock-down cells……..161

5.6 Expression profiling of *CDKAL1* knock-down cells……...164

5.7 Follow-up of differentially expressed genes……....172

5.8 Discussion……180

Chapter 6: Discussion……..184

References……189

Publications arising from this thesis

List of figures

Chapter 1: General introduction

Figure 1.1.2.2.1: Antigen processing and presentation……11

Figure 1.1.2.2.1.1: Differentiation of Th cell subsets……..12

Figure 1.2.1.1: The structure of skin……15

Figure 1.3.1.1: Anatomy and histology of the gastrointestinal tract……18

Figure 1.5.1.1: Clinical presentation of psoriasis……28

Figure 1.5.2.1: Overview of the events leading to the onset of psoriasis……31

Figure 1.5.3.1.1: Graphical representation of the *PSORS1* locus……34

Figure 1.6.1.1: Endoscopic image of CD……42

Figure 1.6.2.1: Overview of the events leading to the onset of CD……44

Chapter 3: Genetic analysis of 26 Crohn's disease (CD) susceptibility loci in psoriasis

Figure 3.2.1: Overview of linkage disequilibrium conservation around SNP rs6908425……..104

Figure 3.2.2: Overview of linkage disequilibrium conservation around SNP rs2836754……..105

Figure 3.2.3: Overview of linkage disequilibrium conservation around SNP rs762421……107

Figure 3.3.1: Position of linkage disequilibrium blocks around SNP rs2836754……...111

Figure 3.4.1: Overview of linkage disequilibrium conservation across the *CDKAL1* gene……114

Figure 3.5.1: Overview of linkage disequilibrium conservation across the *CDKAL1*

locus……...118

Chapter 4: Analysis of *CDKAL1* gene expression

Figure 4.1.1: Graphic representation of *CDKAL1* isoforms……..123

Figure 4.2.1: Electrophoresis analysis of *CDKAL1* RT-PCR products……..125

Figure 4.3.1: Real-time PCR analysis of *CDKAL1* gene expression…….128

Figure 4.3.2: Real-time PCR analysis of *CDKAL1* expression in primary keratinocytes……..130

Figure 4.3.3: Real-time PCR analysis of *CDKAL1* expression in resting and activated lymphocytes……...131

Chapter 5: Functional analysis of the CDKAL1 protein

Figure 5.1.1: Graphic representation of CDKAL1 domains…….136

Figure 5.1.2: Comparative analysis of the CDKAL1 protein sequence……138

Figure 5.1.3: CDKAL1 amino acid sequence variation……139

Figure 5.2.1: CDKAL1 western blot analysis……...142

Figure 5.2.2: GST-CDKAL1 sequence analysis……143

Figure 5.2.3: Induction of GST-CDKAL1 expression……...144

Figure 5.2.4: Western blot analysis of IPTG induction…….145

Figure 5.2.5: α-CDKAL1 antibody specificity characterization……...146

Figure 5.3.1: CDKAL1 is expressed in the cytoplasmic cell fraction……...148

Figure 5.3.2: CDKAL1 localizes to the cell cytoplasm……149

Figure 5.4.1: HEK293T cell line transfection efficiency……...153

Figure 5.4.2: HaCaT cell line transduction efficiency……...154

Figure 5.4.3: Jurkat cell line transduction efficiency……155

Figure 5.4.4: Real-time PCR analysis of *CDKAL1* gene expression in stably transduced HaCaT and Jurkat cell lines……...156

Figure 5.4.5: Clonal expansion of transduced HaCaT cells……..157

Figure 5.4.6: Real-time PCR analysis of *CDKAL1* gene expression in clones derived from stably transduced HaCaT and Jurkat cell lines…….160

Figure 5.5.1: *CDKAL1* knock-down affects cell proliferation……..162

Figure 5.5.2: *CDKAL1* knock-down influences the rate of cell cycle progression…...163

Figure 5.6.1: Graphic representation of the fold change log ratios for the 276 differentially expressed genes detected by transcription profiling of *CDKAL1* knock-down clones……...166

Figure 5.6.2: GeneGo pathway analysis of the differentially expressed genes detected by expression profiling of *CDKAL1* knock-down clones……...168

Figure 5.6.3: Schematic representation of the process leading to the selection of 14 genes to be followed-up…….169

Figure 5.7.1: Follow-up of the differentially expressed genes identified by microarray analysis (HaCaT cells)……..175

Figure 5.7.2: Follow-up of the differentially expressed genes identified by microarray analysis (Jurkat cells).……....178

Figure 5.7.3: Hierarchical clustering analysis of CASP array data for the 13 differentially expressed genes……179

List of tables

Chapter 1: General introduction

Table 1.1.1: Properties of innate and adaptive immune responses……6

Table 1.5.3.2.1: Loci associated with psoriasis susceptibility…….36

Table 1.5.3.2.2: Pleiotropic disease susceptibility loci associated with psoriasis……...39

Table 1.6.1.1: Contrasting features of the two forms of IBD……..41

Table 1.6.3.2.1: Loci associated with CD susceptibility…….....47

Chapter 2: Materials and methods

Table 2.3.2.1.1: Primer sequences……..…….71

Table 2.7.2.1: DNA restriction…….81

Table 2.9.5.1: Separating gel components……...87

Table 2.9.5.2: 5% Stacking gel components……87

Table 2.9.5.3: Antibodies used in western blot experiments……...88

Table 2.11.3.1: TaqMan assays used in this study……...94

Chapter 3: Genetic analysis of 26 Crohn's disease (CD) susceptibility loci in psoriasis

Table 3.1.1: Power calculations……...99

Table 3.1.2: Association analysis results……...101

Table 3.1.3: Associations results for SNPs generating a FDR < 0.01……...102

Table 3.3.1: Analysis of psoriasis associated SNPs in the CASP dataset……..110

Table 3.4.1: Power calculations for SNPs rs714831, rs714830, rs4454125 and rs775202……115

Table 3.4.2: Analysis of allelic heterogeneity at the *CDKAL1* locus…….……115

Table 3.5.1: Power calculations for SNP rs10946398……...119

Chapter 5: Functional analysis of the CDKAL1 protein

Table 5.4.1: Short hairpin sequences used to knock-down *CDKAL1* expression……..152

Table 5.4.2: sh-*CDKAL1* HaCaT clones……158

Table 5.4.3: sh-*CDKAL1* Jurkat clones……..159

Table 5.6.1: Differentially expressed genes selected for follow-up……..170

Abbreviations

°C: Centigrade

Aa: Aminoacid

ATP: Adenosine triphosphate

bp: Base pair

BCA: Bicinchoninic Acid

BSA: Bovine Serum Albumin

Ca^{2+}: Calcium

CARDs: Capase Recruitment Domains

CASP: Collaborative Association Study of Psoriasis

CD: Crohn's disease

cDNA: Complementary Deoxyribonucleic Acid

CDKAL1: CDK5 regulatory subunit associated protein 1-like 1

CDK5RAP1: CDK5 regulatory subunit associated protein 1

chr: Chromosome

CI: Confidence Interval

CIP: Calf Intestinal alkaline Phosphatase

CNV: Copy-Number Variation

CO_2: Carbon dioxide

Ct: Threshold cycle

CVS: Chorionic Villus Sample

DAPI: 4',6-diamidino-2-phenylindole

DCs: Dendritic cells

DDC: Dermal dendritic cells

dH_2O: Distilled water

DMEM: Dulbecc's Modified Eagle's medium

DMSO: Dimethyl Sulfoxide

dMΦ: Dermal macrophages

DNA: Desoxyribonucleic Acid

dNTP: Deoxyribonucleotide Triphosphate

ECL: Enhanced chemiluminescence

EDTA: Ethylenediaminetetraacetic Acid

ER: Endoplasmic Reticulum

Ex: Exon

FCS: Fetal Calf Serum

FDR: False Discovery Rate

FITC: Fluorescein Isothiocyanate

FSC: Forward Scatter

GI: Gastro intestinal

GO: Gene Ontology

GST: Glutathione-S-Transferase

GWASs: Genome-Wide Association Studies

h: Hour

H_2O_2: Hydrogen peroxide

HCl: Hydrochloric acid

HEK293T: Human Embryonic Kidney 293 cells

HRP: Horseradish peroxidase

HWE: Hardy-Weinberg Equilibrium

IBD: Inflammatory Bowel Disease

IFNs: Interferons

IFN-γ: Interferon Gamma

Ig: Immunoglobulin

IL-: Interleukin

IPTG: Isopropyl-β-D-1-thiogalactopyranoside

KCs: Keratinocytes

kDa: KiloDalton

LB: Luria broth

LCs: Langherhans cells

LD: Linkage Disequilibrium

LRR: Leucine-Rich Repeats

LUA: Luria broth (LB)/agar

M cells: Microfold cells

MAPK: Mitogen-Activated Protein Kinase

MDA-5: Melanoma Differentiation-Associated gene 5

mDC: Dermal Myeloid Dendritic cells

mg: Milligram

$MgCl_2$: Magnesium chloride

MGB: Minor Grove Binder

MHC: Major Histocompatibility Complex

min: Minute

mL: Millilitre

mm^3: millimetre cubic

mRNA: Messenger RNA

MTTase: methylthiotransferase n: Number

NaCl: Sodium chloride

NALP: NACHT, LRR and PYD containing proteins

NF-κB: Nuclear Factor kappa-light-chain-enhancer of activated B cells

ng: Nanogram

NH_4SO_4: Ammonium sulfate

NK: Natural Killer

NLRs: NOD-Like Receptors

NOD: Nucleotide-Binding Oligomerization Domain

NTC: No Template Control

OR: Odds ratio

PAMPs: Pathogen-Associated Molecular Patterns

PBS: Phosphate Buffer Saline

PCR: Polymerase chain reaction

pDCs: Plasmacytoid Dendritic cells

PEI: Polyethylenimine

pH: potential of Hydrogen

PHA: Pytohemagglutinin-M

PI: Propidium iodide

PMSF: Phenylmethanesulfonylfluoride

PPIA: Cyclophilin A

PRRs: Pattern Recognition Receptors

PSORS1: Psoriasis susceptibility 1

PYD: Effector Pyrin Domain

RIG-I: Retinoic acid-Inducible Gene-I

RLR: RIG-I like receptor

RMA: Robust Multichip Average

RNA: Ribonucleic Acid

RNAi: RNA interference

RPLPO: Large PO ribosomal protein

RPMI 1640: Roswell Park Memorial Institute medium

RQ: Relative Quantification

RT-PCR: Reverse Transcription-Polymerase Chain Reaction

SAM: radical-*S*-adenosylmethionine

sec: Seconds

sh-RNA: Short hairpin RNA

SNPs: Single Nucleotide Polymorphisms

SSC: Side Scatter

T2D: Type 2 diabetes

TAP: Transporter Associated with Antigen Processing

TB: Transformation buffer

TBE: TRis-Borate-EDTA buffer

TBS: Tris Buffered Saline

Tc: cytotoxic T lymphocytes

TCR: T-cell receptor

TE: Tris-EDTA buffer

TGF: Transforming Growth Factor

tGFP: turbo Green Fluorescent Protein

Th: T helper cells

TIR: Toll-Interlukin-1 Receptor

TLRs: Toll-Like Receptors

TNF-α: Tumor Necrosis Factor Alpha

Treg: Regulatory T cells

tRNA: Transport RNA

U/μL: Unit per microliter

US: United States

UTR: Untranslated region

UV: Ultraviolet

UK: United Kingdom

v/v: Volume/Volume

V: Volt

VEGF: Vascular Endothelial Growth Factor

w/v: Weight/Volume

WTCCC: Wellcome Trust Case-Control Consortium

μL: Microliter

μM: Micromolar

Chapter 1: General introduction

1.1 The immune system

The immune system protects the organism against the attack of pathogens through a variety of recognition systems and effector mechanisms that seek out and destroy invading microorganisms. This is achieved by means of two main types of immune responses known as innate and adaptive immunity (Table 1.1.1).

1.1.1 Innate immunity

Innate immunity is the first response of the immune system against any microbial attack. It is an evolutionary conserved mechanism of defence against invading microorganisms, which is also found in plants, insects, fungi and in primitive multicellular organisms (Beck and Habicht 1996).

1.1.1.1 Innate immune receptors

The innate immune system senses the presence of phatogenes by means of specialized receptors, which recognize microbe-specific molecules. These sensors are known as pattern recognition receptors (PRRs) and their ligands are referred to as pathogen-associated molecular patterns (PAMPs). PRRs can be divided into three main families: Toll-like receptors, RIG-I-like receptors and NOD-like receptors.

1.1.1.1.1 Toll-like receptors

Toll-like receptors (TLRs) are evolutionary conserved membrane-bound proteins located on the cell surface (TLR- 1, 2, 4, 5, 6 and 10), or in endosomes (TLR- 3, 7, 8 and 9). They are formed by a ligand-binding domain containing 19-25 consecutive leucine-rich repeats (LRR), a membrane spanning domain and a conserved signalling

motif, known as the Toll-interlukin-1 receptor (TIR) domain. TLRs recognize evolutionary conserved features present on the cell walls of pathogens (*e.g*. lipoproteins and peptidoglycan), as well as viral nucleic acids. Following ligand recognition, TLRs activate two different cascades known as the MyD88-dependent and the TRIF-dependent pathway. The former results in the activation of the pro-inflammatory transcription factor NF-κB, whereas the latter leads to the production of type I interferons (IFNs) (Kawai and Akira 2008; Fukata, Vamadevan et al. 2009).

1.1.1.1.2 RIG-I like receptors

RIG-I-like receptors RLRs form a second group of PRRs that are able to recognize viral nucleic acids. RLRs reside in the cytoplasm and include the RIG-I (retinoic acid-inducible gene-I) and MDA-5 (melanoma differentiation-associated gene-5) receptors. Both share the same molecular structure, including two amino-terminal Capase Recruitment Domains (CARDs), a central ATPase and helicase domain, and a carboxy C-terminal regulatory motif. Upon binding of viral RNA, RLRs signal through the IPS-1 adaptor and induce the production of type I IFNs, through the activation of IRF3 and NF-κB (Kawai and Akira 2008; Nakhaei, Genin et al. 2009).

1.1.1.1.3 NOD-like receptors

NOD-like receptors (NLRs) are soluble cytosolic proteins, which recognise pathogens and molecules released during cell damage or stress. The 23 NLRs that are encoded by the human genome can be classified in two main subfamilies: NOD (nucleotide-binding oligomerization domain) and NALP (NACHT, LRR and PYD containing proteins) proteins. These share two common domains: the ligand-sensing LRRs and the NACTH (named after the NAIP, CIITA, HET-E and TP-1 proteins, where it is present)

oligomerization motif. However, NOD and NALP proteins differ in the make-up of their effector regions, which consist of a CARD and a PYD (effector pyrin domain) domain, respectively. Following ligand recognition, NOD receptors activate NF-κB and mitogen-activated protein kinase (MAPK) signalling, whereas NALPs cause the cleavage of the pro-inflammatory precursors pro-IL-1β and pro-IL-18 into their mature forms, through activation of caspase 1 and 5 (Fukata, Vamadevan et al. 2009).

1.1.1.2 Innate immune cells

1.1.1.2.1 Keratinocytes

The first line of defence against invading pathogens is provided by anatomical barriers, such as the skin. Keratinocytes (KCs), which are the predominant cell type in the epidermis (*i.e.* the skin upper layer), not only contribute to the integrity of the skin barrier but also participate in innate immune responses. In fact, KCs express most TLRs and RLRs, as well as a variety of NLRs (Miller and Modlin 2007; Miller and Modlin 2007; Harder and Nunez 2009). Upon PAMP recognition, these receptors trigger the production of anti-microbial peptides (*e.g.* β-defensins and chatelicidins) and pro-inflammatory mediators (*e.g.* cytokines and chemokines) (Nestle, Di Meglio et al. 2009).

1.1.1.2.2 Macrophages and neutrophils

Macrophages and neutrophils are two early effectors of the innate immune response, which are stimulated through the activation of their TLRs.

Macrophages are mononuclear white cells resident within tissues. They differentiate from monocytes, which are their circulating progenitors and are found at a concentration of 500 to 1,000 per mm^3 of blood.

Neutrophils are part of the granulocyte family of polymorphonuclear leukocytes. These

are very abundant in the blood as they are present at a concentration of 4,000 to 10,000 per mm^3. This number can rapidly increase to 20,000 per mm^3 in the presence of an infection. This cell type can easily invade extra-vascular tissues at infection sites by binding to endothelial adhesion molecules such as selectin, integrin, ICAM-1 and ICAM-2 and responding to chemoattractants (Lawrence and Springer 1991; Teng, Johkura et al. 2003).

Both macrophages and neutrophils are phagocytes, *i.e.* cells that engulf microbes by extending and wrapping part of their plasma membrane around invaders. Once the microbes are internalized, they are attacked by several enzymes (*e.g*. phagocyte oxidase, inducible nitric oxide synthase and lysosomal proteases) activated within lysosomes and phagolysosomes.

1.1.1.2.3 Natural Killer cells

Natural Killer (NK) cells are a small population of immune cells, which account for 5-10% of circulating lymphocytes and which represent the first line of defence against viruses. NK cells have cytolytic activity towards infected cells, which they recognize, based on abnormalities in the profile of surface antigens. Moreover, activated NK cells produce a variety of cytokines, most notably TNF-α and IFN-γ. These two cytokines promote dendritic cell maturation with a consequent enhancement of antigen presentation to T cells (see section 1.1.2.2). IFN-γ has also a powerful effect on macrophage activation. Thus, NK cells contribute to shaping the overall host immune response to viruses (Vivier, Raulet et al. 2011).

1.1.1.2.4 Dendritic cells

Dendritic cells (DCs) are the professional antigen presenting cells (APC). They continuously circulate between the bloodstream and the tissues that are most exposed to the external environment, *i.e.* the skin and the epithelia lining the lung and the gastro-intestinal tract. Several types of DCs are known. Langherhans cells (LCs), which express the CD1a and CD207 (langerin) surface markers, are found in the epidermis and in lymph nodes. Although LCs have previously been considered to be antigen presenting cells, more recent work suggest that they may also act as immune modulators *in-vivo* (Kaplan, Jenison et al. 2005).

Dermal myeloid dendritic cells (mDCs), which express the surface marker CD1c, differentiate from circulating monocytes that are recruited to the dermis under inflammatory conditions. mDCs are antigen presenting cells, which also contribute to the fight against infection, through the production of cytokines and chemokines (Auffray, Sieweke et al. 2009).

Plasmacytoid DCs (pDCs), which express the BDCA-2 surface marker, differentiate from plasmacytoid precursors. pDCs produce high amounts of IFN-α upon exposure to viral nucleic acids and activation of TLR7 and TLR9 (Nestle, Conrad et al. 2005).

Table 1.1.1: Properties of innate and adaptive immune responses

	Innate response	**Adaptive response**
Recognition mechanism	Direct	Indirect (mediated by antigen presentation)
Specificity	Limited	High
Onset of response	Immediate (hours)	Delayed (days)
Immunological memory	Absent	Present
Self-recognition	Absent	Rare
Clonal expansion	Absent	Present
Organisms	Nearly all multi-cellular organisms	Vertebrate animals only

Adapted from (Mayer 2006)

1.1.2 Adaptive immunity

The adaptive immune system provides a response against pathogens when innate mechanisms are no longer able to contain an infection. Its action is based on antigen-specific recognition of invaders and is also characterized by the key features of memory and delayed onset. These properties distinguish it from innate immunity (Table 1.1.1).

Acquired immunity is dependent on a highly specialized set of cells known as lymphocytes. These are an extremely heterogeneous population divided in two different groups, T and B cells which respectively differentiate from their precursors, in the thymus and in the bone marrow.

1.1.2.1 B lymphocytes and the humoral response

B cells are key players of the humoral response, *i.e.* the production of antibodies mediating the destruction of extracellular microorganisms. Immature B lymphocytes are activated into proliferating and antibody-secreting plasma cells, through their exposure to antigens in lymphoid secondary organs (most notably the spleen and lymph nodes). B cell activation leads to a change in the heavy chain class of the antibody that is produced, a process known as class switching. B cell activation can also result in the phenomenon of affinity maturation, whereby the affinity of the antibody for its antigen increases progressively.

The antibodies secreted by plasma cells and memory B cells have three main effector functions. First, antibodies promote opsonisation, a process whereby phagocytes expressing the FcR heavy chain receptor bind the constant region of Ig molecules and internalize pathogens. Secondly, antibodies activate the complement pathway, a highly regulated cascade of proteolytic cleavage reactions, which results in the lysis of foreign cells. Finally, the interaction of antibodies with Fc receptors expressed on a variety of

effector cell types can trigger their cytotoxic activity, in a process known as antibody-dependent cell-mediated cytotoxicity (Manabu 2010).

1.1.2.2 T lymphocytes and antigen presentation

T cells can be classified into two main subsets, which can be distinguished on the basis of the surface markers they express. $CD4^+$ T helper cells (Th) are mainly involved in the activation of other immune cell types, whereas $CD8^+$ cytotoxic T lymphocytes (Tc) have direct cytolytic activity towards infected cells. A third lymphocyte population is represented by $CD4^+CD25^+$ regulatory T cells (Treg), which suppress the activation of the immune system.

T lymphocytes can only participate to immune responses through the process of antigen presentation, whereby peptides bound to Major Histocompatibility Complex (MHC) proteins are presented to T the cell receptor by macrophages, DCs and B cells. In general, class I MHC molecules, which are expressed on most nucleated cells, participate to the presentation of endogenous antigens to $CD8^+$ Tc cells (Banchereau, Briere et al. 2000; Hansen and Bouvier 2009). Conversely, class II MHC molecules, which are restricted to B cells, macrophages and DCs, participate to the presentation of exogenous antigens to $CD4^+$ Th cells (Banchereau, Briere et al. 2000) (Fig. 1.1.2.2.1).

1.1.2.2.1 T helper cell subsets

Activation of Th cells through antigen presentation generates three different effector populations, known as Th1, Th2 and Th17 (Fig.1.1.2.2.1.1).

Th1 cells are major producers of IFN-γ, which stimulates the activity of macrophages and Tc cells. Th2 cells produce IL-4, which stimulates class-switching in B lymphocytes. Th2 cells also release IL-13 and IL-5, which respectively attract basophils

and eosinophils (Moser and Murphy 2000; Eisenbarth, Piggott et al. 2003).

Th17 cells are the main producers of IL-17 and IL-22, two cytokines which drive the recruitment of neutrophils to sites of inflammation (Dong 2008; Martinez, Nurieva et al. 2008; Nestle, Di Meglio et al. 2009).

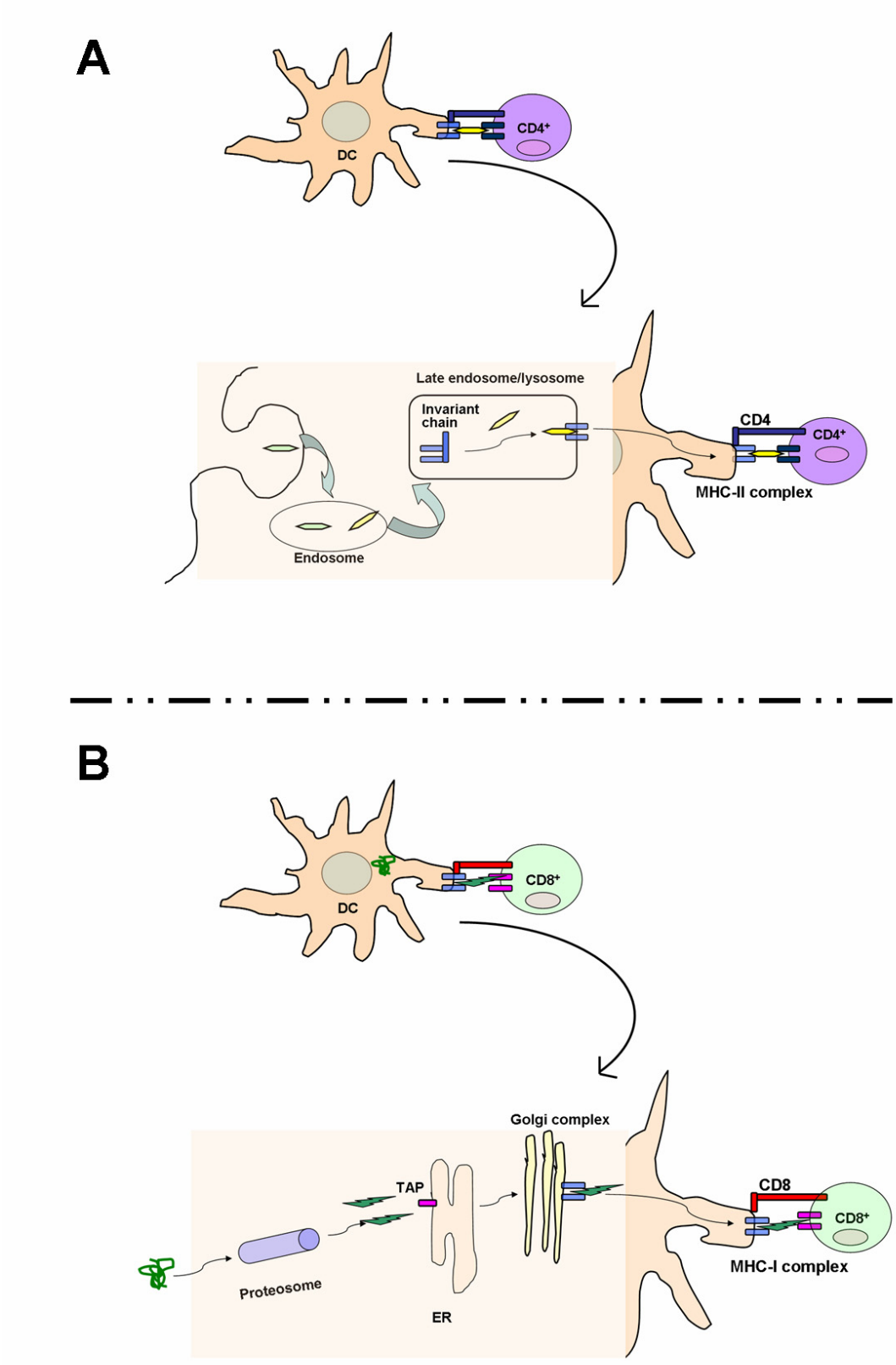

A
DC
CD4+
Late endosome/lysosome
Invariant chain
CD4
CD4+
MHC-II complex
Endosome
B
DC
CD8+
Golgi complex
TAP
CD8
CD8+
MHC-I complex
Proteosome
ER

Figure 1.1.2.2.1: Antigen processing and presentation

(A) Antigen presentation *via* MHC class II molecules. Exogenous antigens are initially degraded in the endosome and then directed into the late endosomal/lysosomal compartment, where they interact with MHC molecules. The peptide-MHC complex is next transported to the cell surface, and presented to the T-Cell Receptor (TCR). This interaction is further stabilized by the presence of the CD4 molecule.

(B) Antigen presentation *via* MHC class I molecules. Endogenous antigens are degraded by the proteasome and the resulting peptides are delivered to the endoplasmic reticulum (ER) by transporter associated with antigen processing (TAP). The peptide-MHC complex which forms in the ER is transported to the cell surface *via* the Golgi apparatus. The antigen is presented to the TCR. This interaction is further stabilized by the CD8 molecule.

Adapted from (Banchereau, Briere et al. 2000)

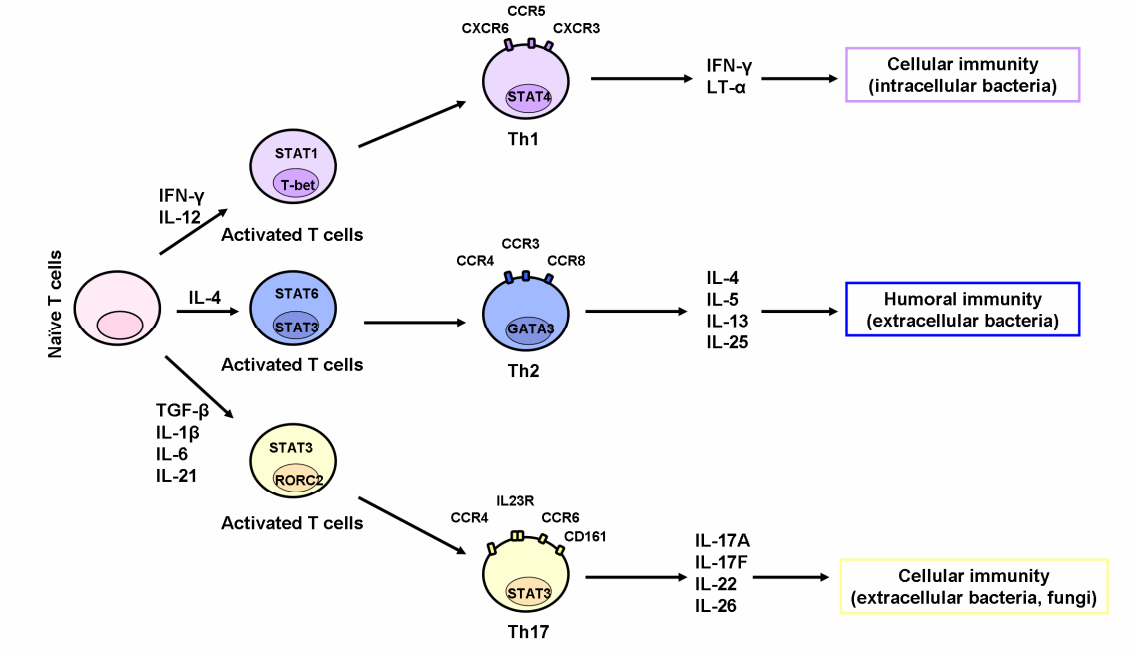


Figure 1.1.2.2.1.1: Differentiation of Th cell subsets

Naïve T cells differentiate into distinct effector subsets, in the presence of different cytokines. Differentiation to Th1 status occurs in the presence of IL-12 and IFN-γ, through the activation of the transcription factor T-bet. Th1 cells express the surface chemokine receptors CXCR6, CXCR5, and CCR3. Th2 differentiation is dependent on the presence of IL-4 and the activation of the STAT6 and STAT3 transcription factors. Committed Th2 cells express CCR4, CXCR3 and CCR8.

Th17 differentiation requires the action of TGF-β1, IL-1β, IL-6 and IL-21, as well as the activation of the RORC2 transcription factor. Th17 cells express CCR4, CCR6 and CD161, as well as IL23R.

Adapted from (Dong 2008; Di Cesare, Di Meglio et al. 2009)

1.2 The structure of skin

The skin is the largest organ of the human body and accounts for approximately 15% of its weight. It defends the body against microbial pathogens, as well as chemical and physical insults. The skin is composed of two layers, known as the epidermis and the dermis.

1.2.1 The epidermis

The epidermis (from the Greek *epi*, on top and *derma*, the skin) represents the outermost compartment of the skin. It consists of five strata through which keratinocytes migrate as they differentiate. The *stratum basale* is the bottom layer of the epidermis and is formed by a single row of basal keratinocytes, which constantly proliferate by mitosis. As these cells differentiate, they move into the *stratum spinosum,* which consists of 8-10 layers of cells connected by desmosomes. The subsequent layer is the *stratum granulosum,* which is formed by 2-4 rows of keratinocytes. These cells have a darker cytoplasm due to the presence of keratoyalin granules. The *stratum lucidum* is a thin layer that is only apparent in the soles of the feet and palms of the hands. It is composed by flatted keratinocytes filled with eladin, a substance which is similar to keratin. The terminal differentiation of keratinocytes is completed in the *stratum corneum,* which is composed of anucleated, flattened dead cells, filled with keratin. These are finally shed in the process of skin desquamation (Fig. 1.2.1.1) (Cristopher 2010).

1.2.2 The dermis

The dermis consists of two layers known as *stratum papillare* and *stratum reticulare*. The *stratum papillare* is rich in lymph and blood vessels, sensory nerve endings, hair follicles and glands. It can protrude into the epidermis through finger-like projections known as dermal papillae. The *stratum reticulare* provides mechanical strength and elasticity to the skin. These characteristics are conferred by a network of fibroblasts, collagenous, elastic and reticular fibres (Cristopher 2010).

1.2.3 Skin resident immune cells

As the skin is continuously exposed to environmental pathogens, it hosts a large population of resident immune cells, including Langherans cells (LCs), dermal mDCs, macrophages and T cells.

LCs and dermal mDCs are antigen presenting cells that participate to inflammatory responses through the release of cytokines and chemokines. A subtype of mDC is specialised in the release of TNF-α and inducible nitric oxide synthase. These cells are known as TIP (TNF-α and iNOS producing) and play an important role in the fight against bacterial infections (Serbina, Salazar-Mather et al. 2003).

Dermal macrophages (dMΦ) can be classified into the MΦ1 and MΦ2 sub-populations, which promote Th1 and Th2 responses, respectively (Mantovani, Sica et al. 2004).

Twice as many T cells are present in the skin than in peripheral blood. $CD8^+$ T cells are mainly found in the epidermis and $CD4^+$ T cells in the dermis. These lymphocytes are the body's first line of defence against antigenic challenges and play a key role in maintaining the skin immune homeostasis (Nestle, Di Meglio et al. 2009).

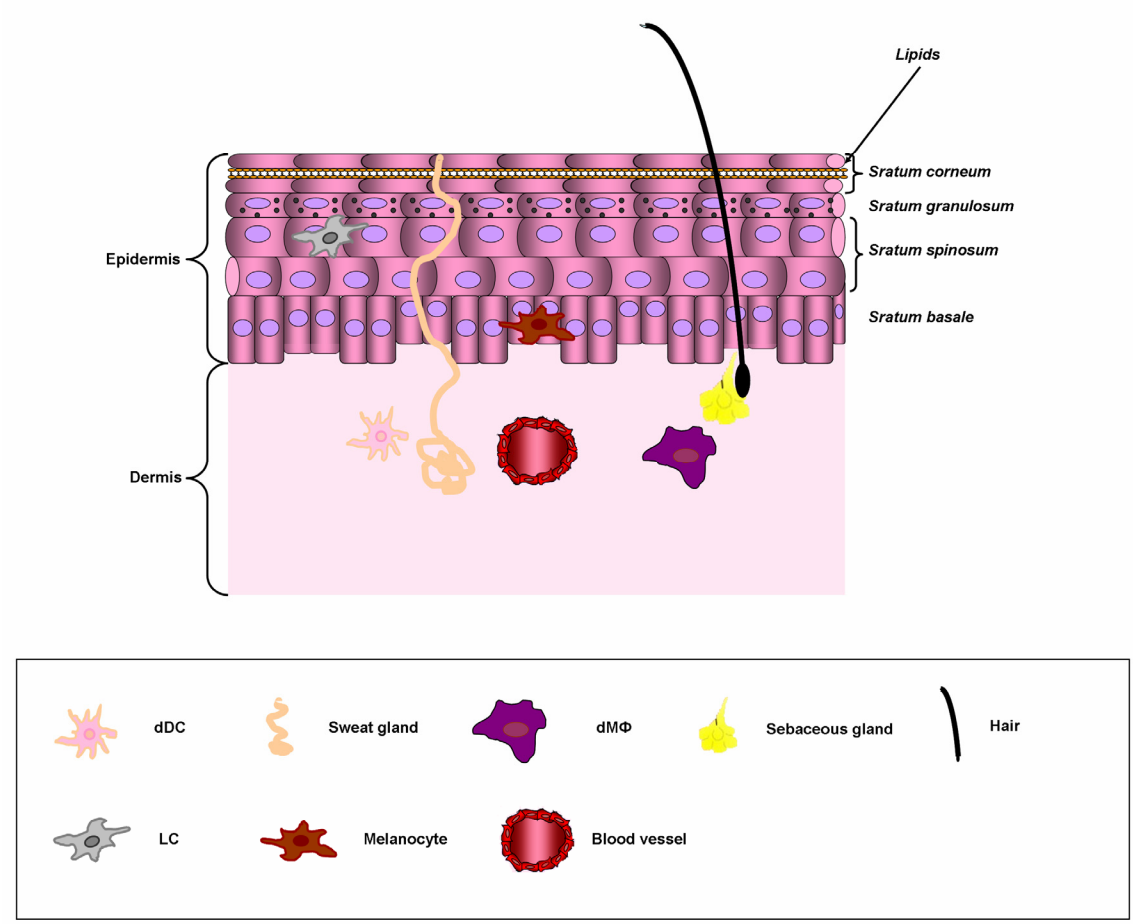


Figure 1.2.1.1: The structure of skin

The dermis and the epidermis host a variety of cell types. The epidermis is mainly composed of keratinocytes, but the *stratum basale* also includes melanocytes, the melanin-producing cells that protect the organism from ultraviolet radiation. Likewise, the *stratum spinosum* contains LCs, as well as keratinocytes. The dermis is mainly composed of fibroblasts, but it also contains macrophages, dermal mDCs, sebaceous and sweat glands, hair follicles and blood vessels.

Adapted from (Nestle, Di Meglio et al. 2009)

1.3 The gastrointestinal mucosa

1.3.1 Histology of the gastrointestinal tract

The gastrointestinal (GI) tract comprises the various organs that extract energy and nutrients from food, while expelling unwanted waste. The upper GI tract includes the oesophagus, stomach and duodenum. The lower GI tract consists of the remaining part of the small intestine (jejunum and ileum) and of the entire large intestine (cecum, colon and rectum).

At the histological level, the GI tract can be divided into four layers known as mucosa, submucosa, *muscularis externa*, and *adventitia*. The innermost layer is the mucosa, which is involved in the digestion and absorption of nutrients. The mucosa consists of three further layers: ephithelium, *lamina propria*, and *muscularis mucosae*. The submucosa is a large, irregular connective tissue layer, characterized by the presence of blood and lymphatic vessels, as well as nerves. The submucosa is also the site of Meissner`s plexus, a network of parasympathetic ganglia that innervates the *muscularis mucosae*. The M*uscularis externa* is composed of two smooth muscle layers (an inner circular one and a longitudinal external one), which are innervated by Auerbach`s plexus. Finally, the adventitia is the external connective tissue the covers the *muscularis externa*. When this layer faces the peritoneal cavity, it becomes covered by mesothelium and takes the name of serosa (Fig. 1.3.1.1) (Burkitt 2000; Cristopher 2010).

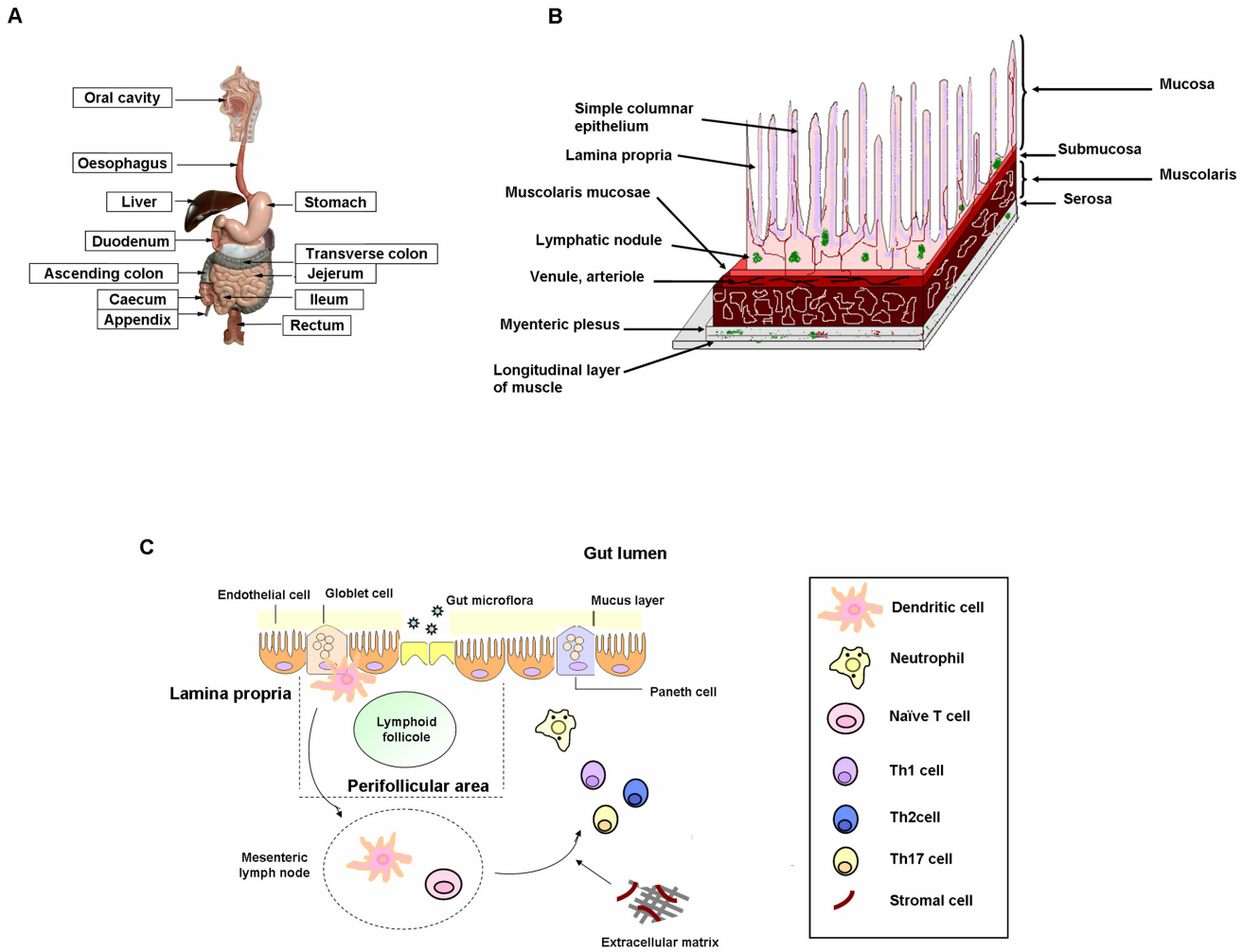

A
Oral cavity
Oesophagus
Liver
Stomach
Duodenum
Transverse colon
Ascending colon
Jejerum
Caecum
Ileum
Appendix
Rectum
B
Simple columnar epithelium
Lamina propria
Muscolaris mucosae
Lymphatic nodule
Venule, arteriole
Myenteric plesus
Longitudinal layer of muscle
Mucosa
Submucosa
Muscolaris
Serosa
C
Gut lumen
Endothelial cell
Globlet cell
Gut microflora
Mucus layer
Paneth cell
Lamina propria
Lymphoid follicole
Perifollicular area
Mesenteric lymph node
Extracellular matrix
Dendritic cell
Neutrophil
Naïve T cell
Th1 cell
Th2cell
Th17 cell
Stromal cell

Figure 1.3.1.1: Anatomy and histology of the gastrointestinal tract

(A) Schematic representation of the gastrointestinal tract (GI) showing the various organs that make up the digestive system. **(B)** Histological structure of the GI tract. The position of the various tissue layers described in section 1.3.1 is shown. **(C)** Key players in the GI immune system. The epithelial layer lining the mucosa contains globet and Paneth cells, as well as Microfold cells (M cells) and DCs which sample the luminal environment for pathogens. Upon exposure to external agents DCs migrate to the mesenteric lymph nodes, where they activate naïve T cells.

Adapted from (Cho 2008)

1.3.2 The mucosal barrier

The epithelium lining the GI tract mucosa forms a barrier which protects the organism against pathogens and dietary antigens, while allowing nutrient absorption, fluid exchange and waste secretion. The first line of defence against microorganisms is provided by mucins, a family of heavily glycosylated proteins that are secreted by specialized goblet cells. Mucins form a hydrated gel that prevents bacteria from coming into contact with the epithelium of the mucosa. Bacteria are also contrasted by Paneth cells, another set of specialized epithelial cells, which fight infections by secreting antimicrobial defensins. A final layer of protection is provided by tight and adherens junctions, which seal the space between cells. Of note, the function of tight junctions can be affected by inflammatory cytokines, such as TNF-α and IFN-γ (Turner 2009).

1.3.3 The mucosal immune system

The defence against infections is of particular importance in the intestinal tract, which is the body's largest mucosal surface. An important role is played by Peyer's patches, which are the specialized lymphoid nodules of the small intestine. These are sites where Microfold cells (M cells) deliver antigens to a variety of DC subsets, *via* transepithelial vesicular transport. Antigen presentation by DCs results in activation of Th1 and Th17 effector cells. T cells that have been activated by DCs can also migrate into the *lamina propria* and the epithelium, where they differentiate into cytotoxic T cells.

The GI tract also hosts a population of IgA-producing plasma cells that reside in the *lamina propria*. The IgA molecules secreted by these cells are transported across the epithelium and into the gut lumen, where they provide another line of defence against pathogens (Strober, Fuss et al. 2002).

1.4 Identification of complex disease susceptibility genes

Genetic factors play an important role in the pathogenesis of most common diseases. In the last few years, the availability of the human genome sequence (Lander, Linton et al. 2001; Venter, Adams et al. 2001) and the advent of high-throughput genotyping technologies have allowed geneticists to make significant progresses in elucidating the inherited basis of disease susceptibility.

1.4.1 The genetic component of complex traits

Assessing the role played by genetic factors is a first, essential step in the analysis of any complex trait. Historically, this has been achieved through epidemiological surveys investigating the familial aggregation of diseases or their concordance rates in monozygotic and dizygotic twins. The results of these studies have informed the calculation of parameters such as the λs (a measure of the increase in disease risk among siblings of affected patients), or h^2 (an index of the proportion of phenotypic variance that can be attributed to genotypic variance), which can be used to quantify the magnitude of the genetic component underlying a given phenotype (Strachan 2004).

1.4.2 Genome-wide linkage studies

A classical procedure for determining the chromosomal location of a disease gene is the linkage study, where the co-segregation of markers and disease phenotypes is assessed across pedigrees. Linkage studies have been used with great success in the genetic analysis of monogenic disorders, but have mostly generated inconclusive results, when applied to non-Mendelian conditions. In fact, linkage methods rely on the accuracy of the inheritance model that is used to analyse the data. When the genetic architecture of a disease is unknown, linkage performs very poorly (Strachan 2004). Moreover, it has

been established that linkage affords very limited power for the detection of common variants of small phenotypic effect (Risch 2000). For these reasons, linkage analysis had very limited success in uncovering complex disease susceptibility alleles, so much so, that the identification of *NOD2* mutations in CD (Hugot, Chamaillard et al. 2001; Ogura, Bonen et al. 2001), is considered one of the very few genuine findings emerging from these studies.

1.4.3 Genome-wide association studies (GWASs)

Association studies are based on the comparison of case and control allele frequencies. Although theoretical calculations had long established that this approach had adequate statistical power for the detection of small phenotypic effects (Risch 2000), the practical implementation of genome-wide association studies (GWASs) had to await the development of platforms for high-throughput genotyping and the availability of sufficient information on the human genome variation content.

1.4.3.1 The HapMap Project

The HapMap project was launched in 2002 with the objective of cataloguing DNA sequence variants in individuals of different ethnicity. The project is being implemented in three sequential phases. Two-hundred and sixty-nine individuals of European, African and Asian descent were analysed in phase I and genotypes were obtained for more than $1x10^6$ Single Nucleotide Polymorphisms (SNPs), occurring with a minor allele frequency > 0.05. These data provided a general picture of Linkage Disequilibrium (LD) conservation across the human genome (The International HapMap Consortium 2005). In phase II, genotyping was extended to include more than 3.1×10^6 SNPs, thus generating a comprehensive and freely accessible genetic resource for the selection of markers to be used in GWASs (Frazer, Ballinger et al. 2007). Finally, in phase III the

study population was extended to include a total of 1,184 individuals, the genotypes of which formed the basis for an integrated dataset, including information on common and rare alleles, as well as copy-number variation (CNV) (Altshuler, Gibbs et al. 2010).

1.4.3.2 The 1,000 Genome Project

The 1,000 Genomes Project was launched in 2008, with the aim of exploiting next-generation sequencing technologies in order to generate a comprehensive catalogue of human genetic variation (including 95% of variants occurring at > 0.01 frequency in non-coding regions and > 0.001 frequency in coding exons), in five major population groups (originating from Europe, East Asia, South Asia, West Africa and the Americas). This objective is being pursued by low coverage sequencing of the entire genome and deep targeted sequencing of all coding regions (The 1000 Genome Project Consortium 2010).

1.4.3.3 Experimental design of GWASs

A typical GWAS consists of several steps, the first of which is the ascertainment of patient and control datasets. In this context, the choice of the controls that will be matched with cases is of great importance, as pairing with controls from different ethnic groups will lead to spurious associations. One of the approaches that are currently being used is the analysis of well characterized, prospectively ascertained resources such as the 1958 British Birth Cohort (Power and Elliott 2006).

Once the cases and controls have been recruited, DNAs are genotyped on commercially available platforms, the design of which may be informed by the nature and position of SNPs (gene-centred approach) or the information generated by the HapMap project (indirect approach). The genotype data are next subject to stringent quality controls (*e.g.*

the analysis of Hardy-Weinberg equilibrium and SNP call rates) to eliminate inaccurate calls. Data can also be derived for markers that have not been experimentally analysed, using imputation methods to predict SNP genotypes based on those of neighbouring markers (Howie, Donnelly et al. 2009). Association tests are next performed and replication of significant findings is sought in an independent resource. Because of the extensive multiple testing carried out in these studies, novel associations can only be declared if they meet the so-called threshold for genome-wide significance ($P < 5 \times 10^{-8}$) (Pearson and Manolio 2008).

Meta-analyses can also be carried out across different GWASs, in order to uncover subtle effects that can only be identified through the analysis of very large samples.

1.4.3.4 Successes and limitations of GWASs

The GWASs carried out in the last few years have identified hundreds of disease susceptibility genes, revolutionizing our understanding of common disease genetics. Despite these huge steps forward, a substantial portion of complex disease heritability remains unaccounted for (Maher 2008). This is thought to reflect, at least in part, the limitations of GWASs, which are typically designed to identify common SNPs and are unlikely to detect the effects of rare variants and CNVs (Maher 2008).

1.4.4 Follow-up of association signals

It is important to emphasize that GWASs identify disease susceptibility intervals that typically include more than one gene. Even when that is not the case, the disease associated SNPs that are detected within a given gene are unlikely to include the causal susceptibility allele. Thus, the results of GWASs need to be followed-up by further genetic and functional studies, aimed at identifying the variants underlying association

signals.

1.4.4.1 Fine mapping of disease susceptibility intervals

The fine mapping of association signals used to require the re-sequencing of large genomic segments, an expensive and labour-intensive step that was necessary to catalogue all the genetic variation found in a given disease susceptibility interval. Thanks to the efforts of the 1,000 Genomes and the HapMap projects, this data is now freely available to the scientific community and is being used for the design of custom genotyping platforms. The most notable example is the Immunochip genotyping array, which includes 200,000 SNPs, spanning 186 distinct loci identified in immune disease GWASs. The Immunochip will be used for the analysis of a variety of case-control datasets and is expected to identify the causal susceptibility alleles underlying conditions such as Crohn's disease, celiac disease, multiple sclerosis, psoriasis, and type 1 diabetes.

1.4.4.2 Functional studies

There have been many instances where GWASs have identified disease susceptibility genes of unknown function. Investigating the physiological and pathological role of these genes remains a lengthy and laborious task, which is often complicated by the necessity of generating and validating the required laboratory reagents. These include cDNA constructs and antibodies, but also stable knock-down cell lines or knockout organisms.

Even when the function of a disease gene is well known, dissecting the impact of the causal susceptibility allele can be challenging. The most notable successes have been obtained in *ex-vivo* studies, where gene-expression levels or immune responses were

compared in healthy individuals bearing different genotypes at the locus of interest (Dendrou, Plagnol et al. 2009; Di Meglio, Di Cesare et al. 2011). This underscores the importance of well characterized healthy donor panels (*e.g.* the Cambridge BioResource, including 9,000 individuals who donated their DNA and agreed to be re-called for further analyses) for the follow-up of GWAS results.

1.5 Psoriasis

1.5.1 Clinical presentation

Psoriasis is a chronic and relapsing immune-mediated skin disease that affects 2-3% of the Caucasian population and 0.4-0.7% of individuals of African and Asian descent (Christophers 2001; Nestle, Kaplan et al. 2009).

The disorder affects both sexes with equal frequency and the onset of symptoms shows a bimodal distribution peaking at 15-30 and 50-60 years of age. In this context, the disease can be triggered by a variety of environmental factors, including exposure to certain drugs (*e.g*. lithium and beta-blockers), streptococcal infections, skin trauma and stress (Griffiths and Barker 2007).

Several subtypes of the disease have been described, including psoriasis vulgaris, as well as erytrodermic, guttate, palmoplantar and nail psoriasis.The most common form of the disease (affecting ~ 90% of patients) is psoriasis vulgaris, which is characterised by the occurrence of red plaques that may vary in size and yet are well delimited from the normal skin. Generally, these lesions are localized on elbows, knees, scalp, lumbosacral region, trunk and lower extremities (Fig. 1.5.1.1) (Griffiths, Christophers et al. 2007).

The main histological features of psoriatic plaques are epidermal hyperplasia, angiogenesis, altered differentiation of keratinocytes and infiltration of mononuclear leukocytes in the dermis and the epidermis (Griffiths and Barker 2007).

Whilst rarely fatal, the disorder has a profound impact on patient well-being, with substantial psychological consequences. Several studies have also reported an association with a number of significant co-morbidities, including psoriatic arthritis (a potentially disabling form of joint inflammation), Crohn's disease (CD), myocardial infarction (MI), type 2 diabetes (T2D), dyslipidemia and obesity (Najarian and Gottlieb

2003; Wolters 2005; Gelfand, Neimann et al. 2006; Shapiro, Cohen et al. 2007). Different therapeutic approaches are now available for the treatment of psoriasis. These include the use of topical (*e.g.* corticosteroids, calcipotriene, and tazarotene), systemic (*e.g.* retinoids, methotrexate and cyclosporine) and biologic agents (*e.g.* anti-TNF antibodies). The choice between these options is dependent on the severity of disease and the occurrence of potential side-effects (Gottlieb, Korman et al. 2008; Menter, Gottlieb et al. 2008). Of note, although the treatments outlined above can improve the condition of patients none can cure the disease (Gottlieb, Korman et al. 2008; Menter, Gottlieb et al. 2008).

Figure 1.5.1.1: Clinical presentation of psoriasis

Psoriatic lesions occur in different shapes and sizes. Lesions, which tend to be symmetric, can cover a substantial portion of the body surface. Nails can also be affected (last photo on the right).

Adapted from (Nestle, Kaplan et al. 2009)

1.5.2 Immunopathogenesis

The main cell types found in psoriatic plaques are KCs, DCs, macrophages and T lymphocytes (Barker, Mitra et al. 1991; Nestle, Turka et al. 1994; Nickoloff 1999).

In psoriatic lesions, hyper-proliferating KCs reach the *stratum corneum* of the epidermis while their differentiation is still incomplete. These cells are activated by infiltrating T lymphocytes and release cytokines, angiogenic factors (*e.g.* vascular endothelial growth factor, VEGF), adhesion molecules (*e.g.* intercellular adhesion molecule 1, ICAM-1), and antimicrobial peptides (*e.g.* LL-37 and defensins). Of note, the antimicrobial peptide LL-37 can bind self-DNA to form aggregate structures that trigger type I IFN production in pDCs, *via* activation of TLR9 (Lande, Gregorio et al. 2007). Given that the release of type I IFN from pDCs is an early pathogenic event in psoriasis (Nestle, Conrad et al. 2005), the loss of innate tolerance to self-DNA is now thought to be one of the drivers of autoimmunity in psoriasis (Lande, Gregorio et al. 2007).

Lesional mDCs produce high levels of IL-2 and IFN-γ, thus contributing to Th1 responses (Chu, Di Meglio et al. 2011). mDCs can also sense LL-37/self-RNA complexes *via* TLR8, thus driving the production of the pro-inflammatory cytokines TNF-α and IL-6 (Ganguly, Chamilos et al. 2009).

Macrophages are found in increased numbers in psoriatic skin, especially at the epidermal/dermal interface (Clark and Kupper 2006). Studies carried out in mouse models of the disease indicate that these cells are a significant source of TNF-α (Clark and Kupper 2006; Stratis, Pasparakis et al. 2006; Wang, Peters et al. 2006). These studies have also shown that the activation of macrophages requires the presence of T cells, an observation that is in agreement with the widespread view that psoriasis is a T cell-mediated disease. In fact, T cells numbers are dramatically increased in psoriatic skin, with $CD4^+$ and $CD8^+$ lymphocytes infiltrating the dermis and the epidermis,

respectively. Th1 and Th17 cell numbers are especially elevated in psoriatic plaques and the cytokines they produce are over-expressed (Kagami, Rizzo et al. 2010).

Of note, the cytokines released by Th17 cells promote keratinocyte proliferation as well as production of microbial peptides (Liang, Tan et al. 2006). Thus the interaction between T cells and KCs is crucial to the onset and maintainance of chronic skin inflammation (Fig. 1.5.2.1).

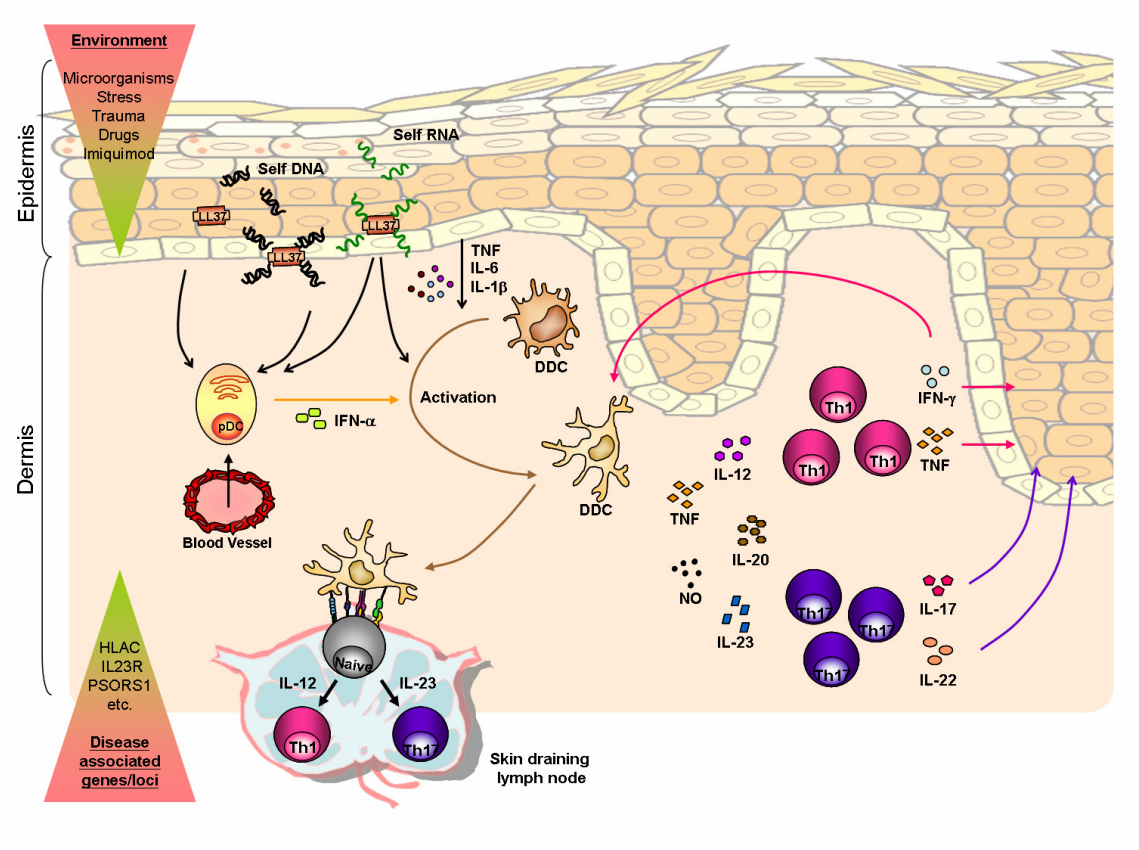


Figure 1.5.2.1: Overview of the events leading to the onset of psoriasis

In psoriatic skin, inflammation is mediated by a crosstalk between cells of the innate and adaptive immune system. The cytokines released by activated KCs (TNF-α, IL-6, IL-1β) and pDCs (IFN-α) lead to the maturation of dermal dendritic cells (DDC). These migrate to lymph nodes, where they promote the differentiation of Th1 and Th17 cells. The cytokines produced by these activated lymphocytes (IFN-γ, IL-17, IL-22) stimulate KC activation, in a positive feedback look that amplifies and sustains chronic skin inflammation.

Adapted from (Chu, Di Meglio et al. 2011)

1.5.3 Genetics of psoriasis

The existence of a genetic component for psoriasis is supported by repeated observations of family disease clustering and higher disease concordance in monozygotic compared to dizygotic twins (Brandrup, Holm et al. 1982; Roberson and Bowcock 2010). On this basis, psoriasis has long been considered as a multifactorial disease caused by gene-gene and gene-environment interactions.

1.5.3.1 The *PSORS1* locus

In the last 15 years, linkage and association studies have repeatedly identified a major disease susceptibility locus, known as *PSORS1* (Psoriasis Susceptibility 1) and lying with the Major Histocompatibility Complex (MHC) on chromosome 6p21.3 (Nair, Henseler et al. 1997; Trembath, Clough et al. 1997; International Psoriasis Genetics Consortium 2003). The susceptibility region has been subsequently refined to a 300 kb DNA segment encompassing part of the MHC class I region (Veal, Capon et al. 2002; Fan, Yang et al. 2008). Several genes lie in this interval, including *HLA-C, PSORS1C3, POU5F1, TCF19, CCHCR1, SPR1, SEEK1, CDSN* and *STG* (Fig. 1.5.3.1.1). Of these, the *HLA-C, CDSN* and *CCHCR1* genes are very polymorphic and carry alleles that are associated with psoriasis across different populations (Tazi Ahnini, Camp et al. 1999; Asumalahti, Laitinen et al. 2000; Enerback, Enlund et al. 2000; Enerback, Nilsson et al. 2000; Asumalahti, Veal et al. 2002; Capon, Allen et al. 2004).

HLA-C encodes a MHC class I protein involved in the presentation of intracellular antigens to $CD8^+$ T lymphocytes (Falk, Rotzschke et al. 1993). *CCHCR1* (alpha-Helix Coiled coil Rod homologue) encodes a ubiquitously expressed protein of unknown function, which is markedly up-regulated in psoriatic skin (Asumalahti, Laitinen et al. 2000). Finally, *CDSN* encodes corneodesmosin, which is a structural protein involved in

keratinocyte adhesion and desquamation (Simon, Jonca et al. 2001).

The strong LD conservation which characterizes the *PSORS1* interval has long confounded the interpretation of these association data (Capon, Munro et al. 2002). However, the results of deep re-sequencing experiments (Nair, Stuart et al. 2006) and GWASs indicate that *HLA-C* is the most likely *PSORS1* candidate gene (Nair, Duffin et al. 2009; Strange, Capon et al. 2010; Sun, Cheng et al. 2010). Of note, the *HLA-Cw*0602* allele has long been associated with psoriasis (Elder, Nair et al. 1994; Mallon, Bunce et al. 1997) but it is still unclear whether it represents the causal disease susceptibility variant or just a marker that is in LD with it.

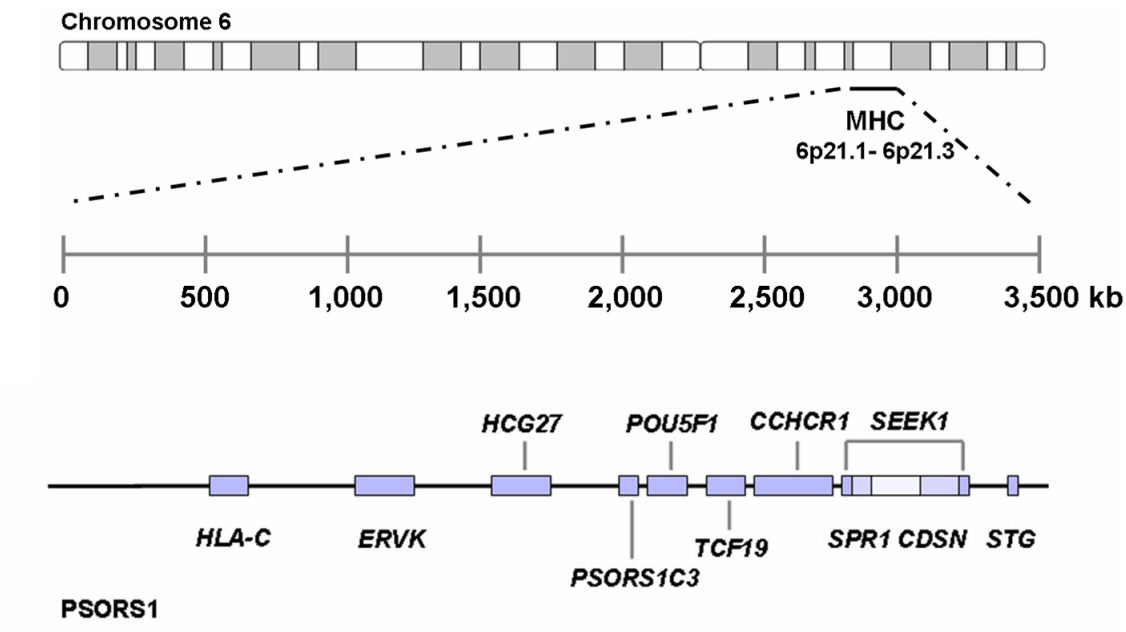


Figure 1.5.3.1.1: Graphical representation of the *PSORS1* locus

The position of the MHC on chromosome 6p21 is shown in the upper panel. The position of known genes within the *PSORS1* interval is shown in the lower part of the figure. The *SEEK1*, *SPR1* and *CDSN* gene regions appear to overlap, as the three genes are transcribed from opposite DNA strands, on a common genomic segment.

Adapted from (Bowcock and Krueger 2005)

1.5.3.2 Non-MHC disease susceptibility loci

Outside of the MHC, the results of genome-wide linkage scans have been inconclusive, reflecting the inadequacy of these studies for the analysis of loci with small phenotypic effects (International Psoriasis Genetics Consortium 2003). Conversely, the advent of GWASs has had a profound effect on our understanding of psoriasis genetics. In fact, the nine GWASs that have been carried out so far have allowed the identification of 23 disease susceptibility loci (Table 1.5.3.2.1). It is interesting to observe that the majority of the genes identified in these studies can be linked to a small number of immune pathways. The most prominent is IL-23/IL-17/NF-κB axis, which plays a key role in the activation of Th17 lymphocytes (Di Cesare, Di Meglio et al. 2009).

Large-scale genetic studies have also identified significant associations between psoriasis and CNVs from the β-defensin and Late Cornified Envelope (*LCE*) gene clusters (Hollox, Huffmeier et al. 2008; de Cid, Riveira-Munoz et al. 2009). These findings indicate that the disruption of keratinocytes innate responses and barrier function also contribute to disease susceptibility.

Of interest, a number of the psoriasis susceptibility variants identified in these studies have also been associated with other immune-mediated conditions (Table 1.5.3.2.2). This phenomenon, that is by no means unique to psoriasis, indicates the existence of pleiotropic genes contributing to the pathogenesis of multiple, clinically distinct diseases.

Table 1.5.3.2.1: Loci associated with psoriasis susceptibility

Gene*	Chr.	Protein function**	Pathway	References
IL23R	1p31	Subunit of the IL-23 receptor	IL-23 signalling	(Cargill, Schrodi et al. 2007; Capon, Bijlmakers et al. 2008; Nair, Duffin et al. 2009; Ellinghaus, Ellinghaus et al. 2010; Strange, Capon et al. 2010)
IL28RA	1p36	Subunit of the IL-29 receptor	IFN signalling	(Strange, Capon et al. 2010)
LCE3B/3C/3D	1q21	Keratinocyte structural protein	Skin barrier function	(de Cid, Riveira-Munoz et al. 2009; Zhang, Huang et al. 2009; Ellinghaus, Ellinghaus et al. 2010)
REL	2p16	Subunit of the NF-κB transcription factor	NF-κB signalling	(Strange, Capon et al. 2010)
IFIH1/ MDA5	2q24	Innate antiviral receptor	IFN signalling	(Strange, Capon et al. 2010)
ERAP1	5q15	Amino peptidase processing MHC class I ligands	Antigen presentation	(Strange, Capon et al. 2010; Sun, Cheng et al. 2010)
IL4, IL13	5q31	Cytokines produced by Th2 cells	IL-4/IL-13 signalling	(Nair, Duffin et al. 2009)
IL12B	5q33	Subunit shared by the IL-12 and IL-23 cytokines	IL-23 signalling	(Cargill, Schrodi et al. 2007; Nair, Duffin et al. 2009; Zhang, Huang et al. 2009; Ellinghaus, Ellinghaus et al. 2010) (Huffmeier, Uebe et al. 2010)***
TNIP1	5q33	Inhibitor of TNF-induced NF-κB-dependent gene expression	NF-κB signalling	(Nair, Duffin et al. 2009; Strange, Capon et al. 2010; Sun, Cheng et al. 2010)
PTTG1	5q33	Anaphase promoting complex substrate	Cell cycle control/DNA repair	(Sun, Cheng et al. 2010)

HLA-C	6p21	MHC class I antigen	Antigen presentation	(Capon, Bijlmakers et al. 2008; Nair, Duffin et al. 2009; Zhang, Huang et al. 2009; Ellinghaus, Ellinghaus et al. 2010; Strange, Capon et al. 2010) (Huffmeier, Uebe et al. 2010)***
TRAF3IP2	6q21	Adaptor mediating IL-17 induced NF-κB activation	IL-17/NF-κB signalling	(Ellinghaus, Ellinghaus et al. 2010; Strange, Capon et al. 2010) (Huffmeier, Uebe et al. 2010)***
TNFAIP3	6q23	Inhibitor of TNF-induced NF-κB activation apoptosis	NF-κB signalling	(Nair, Duffin et al. 2009; Strange, Capon et al. 2010)
CSMD1	8p23	Tumour suppressor gene	Unknown	(Sun, Cheng et al. 2010)
IL23A	12q13	Subunit of the IL-23 cytokine	IL-23 signalling	(Nair, Duffin et al. 2009; Strange, Capon et al. 2010)
GJB2	13q11	Gap junction protein	Electrolyte transport; keratinocyte differentiation	(Sun, Cheng et al. 2010)
NFKBIA	14q13	Inhibitor of NF-κB activation	NF-κB signalling	(Strange, Capon et al. 2010; Stuart, Nair et al. 2010)
FBXL19	16p11	Putative inhibitor of NF-κB activation	NF-κB signalling	(Stuart, Nair et al. 2010)
NOS2	17q11	Induced nitric oxide synthase	Innate antibacterial responses	(Stuart, Nair et al. 2010)
SERPINB8	18q21	Serine protease inhibitor	Unknown	(Sun, Cheng et al. 2010)
TYK2	19p13	Tyrosine kinase associated with cytokine receptors	IL-23 and IFN signalling	(Strange, Capon et al. 2010)
ZNF816A	19q13	Zinc Finger Protein	Unknown	(Sun, Cheng et al. 2010)

ZNF313/RNF114	20q13	E3 ubiquitin ligase	IFN signalling	(Capon, Bijlmakers et al. 2008; Nair, Duffin et al. 2009; Strange, Capon et al. 2010; Stuart, Nair et al. 2010)

*Gene of interest found in the disease-associated susceptibility interval; **Function annotated in the Uniprot protein database; *** Psoriatic arthritis GWAS.

Table 1.5.3.2.2: Pleiotropic disease susceptibility loci associated with psoriasis

Gene	Chr	Other Associated Diseases	References
IL23R	1p31	Crohn's disease; Ankylosing spondylitis	(Duerr, Taylor et al. 2006; Burton, Clayton et al. 2007)
REL	2p16	Rheumatoid arthritis; Celiac disease; Crohn's disease	(Gregersen, Amos et al. 2009; Trynka, Zhernakova et al. 2009; Franke, McGovern et al. 2010)
IFIH1	2q24	Type I diabetes	(Smyth, Cooper et al. 2006; Nejentsev, Walker et al. 2009)
ERAP1	5q15	Ankylosing spondylitis	(Burton, Clayton et al. 2007)
IL12B	5q33	Crohn's disease; Ulcerative colitis	(Parkes, Barrett et al. 2007; Fisher, Tremelling et al. 2008)
PTTG1	5q33	Systemic lupus erythematosus	(Han, Zheng et al. 2009)
TNIP1	5q33	Systemic lupus erythematosus	(Han, Zheng et al. 2009) (Gateva, Sandling et al. 2009)
TNFAIP3	6q23	Rheumatoid arthritis; Systemic lupus erythematosus; Celiac disease	(Plenge, Cotsapas et al. 2007; Thomson, Barton et al. 2007; Graham, Cotsapas et al. 2008; Musone, Taylor et al. 2008; Trynka, Zhernakova et al. 2009)
TYK2	19p13	Type I diabetes; Multiple sclerosis; Systemic lupus erythematosus; Crohn's disease	(Sigurdsson, Nordmark et al. 2005; Australia and New Zealand Multiple Sclerosis Genetics Consortium (ANZgene) 2009; Franke, McGovern et al. 2010; Wallace, Smyth et al. 2010)

1.6 Crohn's disease

1.6.1 Clinical presentation

Crohn's Disease (CD) is a subtype of Inflammatory Bowel Disease (IBD) (Table 1.6.1.1), an idiopathic, chronic-relapsing condition occurring with a higher incidence in the Northern hemisphere (7 per 100,000 in the United States). CD can affect either sex, but an excess of female patients has been reported in all geographical regions (Mathew and Lewis 2004). Several factors can trigger the development of the disease, which usually occurs between the ages of 20 and 40. In fact, smoking, sedentary lifestyle and the alteration of the mucosal intestinal barrier are well recognised risk factors for CD.

Although CD can affect any part of the GI tract, the terminal ileum and the ascending colon are the most frequent sites of inflammation. The distribution of lesions is characteristically patchy, with segments of normal tissue separating affected areas (Fig. 1.6.1.1).

The main symptoms of CD are abdominal pain and diarrhoea, leading to weight loss, tiredness, malabsorption and malnutrition. The GI tract inflammation can also result in bleeding, intestinal obstruction and fistulae formation. Of note, several studies have reported that patients affected by CD are at increased risk of developing other inflammatory diseases, such as primary sclerosing cholangitis, ankylosing spondylitis and psoriasis (Cho 2008).

Different therapeutic approaches are available for the treatment of CD. These include the use of glucocorticoids, immunosuppressive drugs (*e.g.* azathioprine and 6-mercaptopurine) and biologics (*e.g.* infliximab, a chimeric IgG-1 monoclonal antibody against TNF). Although these approaches can relieve disease symptoms, none of them can induce a permanent remission of the pathology.

Table 1.6.1.1: Contrasting features of the two forms of IBD

Features	Crohn's Disease	Ulcerative colitis
Location	Any part of the GI tract	Colon
Lesion distribution	Patchy	Continuous
Nature of lesions	Focal lesions, granulomas	Diffuse lesions
Nature of inflammation	Transmural	Superficial

Adapted from (Satsangi, Silverberg et al. 2006)

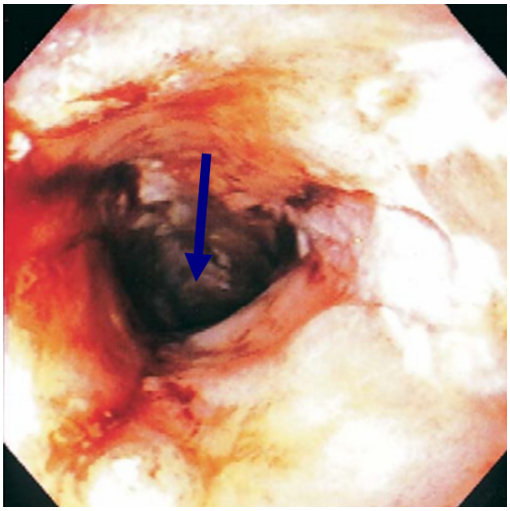


Figure 1.6.1.1: Endoscopic image of CD

An image of a CD patient inflamed bowel. Localized erythema with erosions and stricture of the lumen (blue arrow) are shown.

Adapted from (Hanauer 2009)

1.6.2 Immunopathogenesis

The disruption of GI barrier homeostasis is an early event in CD pathogenesis, which leads to an abnormal activation of innate and adaptive immune responses against resident microbiota. Thus, false recognition of commensal bacteria by DCs and intestinal epithelial cells results in macrophage and NK cell stimulation, as well as Th1 and Th17 cell activation (Dharmani and Chadee 2008) (Fig. 1.6.2.1). The time of Th1 and Th17 involvement is of particular interest. In fact, an experimental tri-nitrophenil (TNP)-colitis mouse model shows a bimodal distribution of cytokine release, characterized by an early Th1 phase, followed by a distinct Th17 response (Strober and Fuss 2011). In the final stages of disease pathogenesis, these lymphocyte populations migrate to the intestinal mucosa, where they stimulate macrophages to release the tissue damaging molecules (*e.g.* nitric oxide, free oxygen radicals and matrix metallo-proteinases) that are responsible for the onset of CD symptoms (Dharmani and Chadee 2008) (Fig. 1.6.2.1).

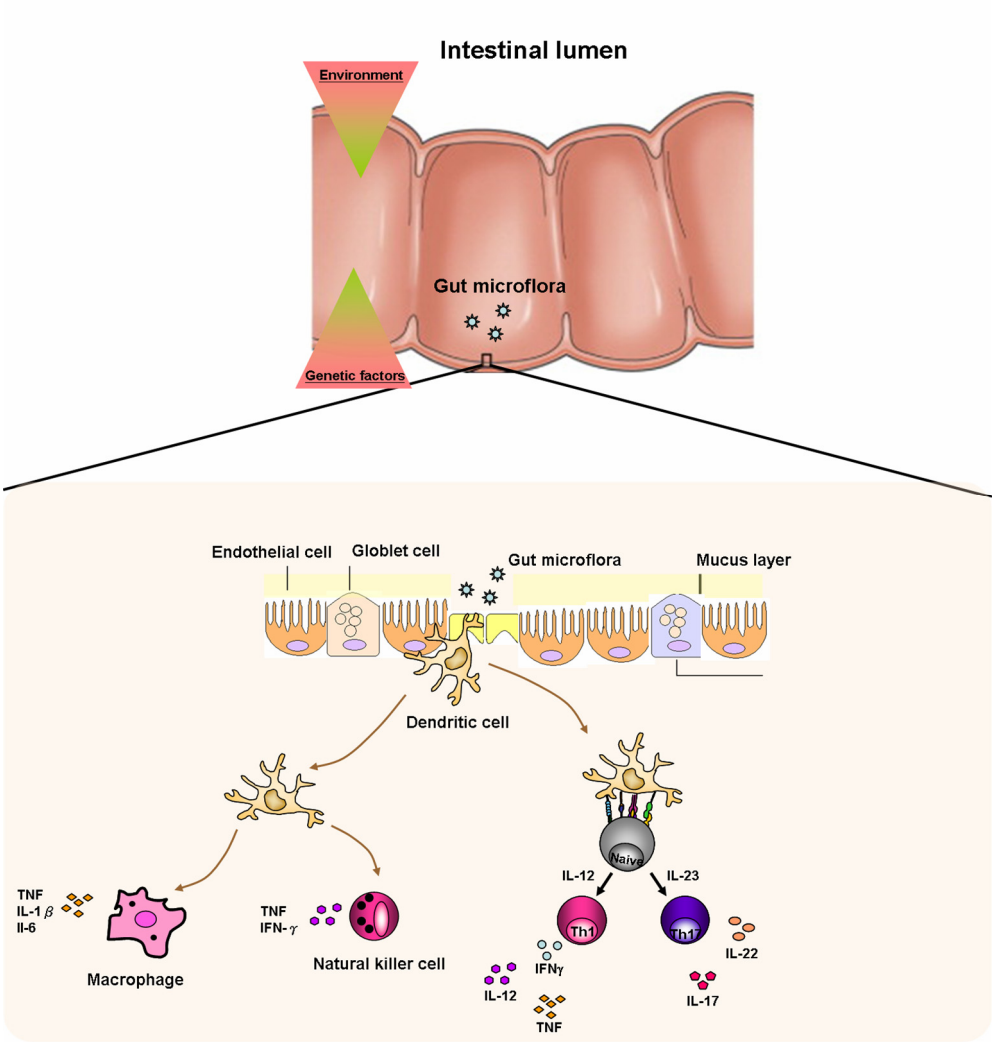


Figure 1.6.2.1: Overview of the events leading to the onset of CD

In CD, inflammation is mediated by a crosstalk between cells of the innate and adaptive immune system. Activation of DCs by the gut microflora causes the polarization of naïve T lymphocytes into Th1 and Th17 cells, as well as the activation of Natural Killer cells and macrophages. The release of cytokines from these cells results in chronic epithelial inflammation.

Adapted from (Strober and Fuss 2011)

1.6.3 Genetics of Crohn's disease

The existence of a genetic component for CD is supported by repeated observations of higher concordance rates in monozygotic (40-60%) compared to dizygotic twins (0-12%) (Podolsky 2002). On this basis, CD has long been considered as a complex disease, caused by gene-gene and gene-environment interactions.

1.6.3.1 The *NOD2/CARD15* susceptibility locus

Genome-wide linkage scans carried out in the late 1990s led to the identification of *NOD2/CARD15*, as the first disease susceptibility gene for CD (Hugot, Chamaillard et al. 2001; Ogura, Bonen et al. 2001). *NOD2* encodes a NOD-like receptor, which drives NF-κB mediated transcription, in response to the muramyl dipeptide constituent of bacterial cell walls (McDonald, Inohara et al. 2005). Approximately 30-50% of CD patients of European descent carry one of three *NOD2* loss-of-function variants (R702W, G908R, L1007fsinsC). These alleles are associated with the strongest increases in disease risk observed for any CD gene, with heterozygous odds ratios ranging from 1.75 to 4 (Abraham and Cho 2009).

1.6.3.2 Other Crohn's disease susceptibility loci

Linkage studies also identified disease susceptibility loci lying within the MHC (van Heel, Fisher et al. 2004) and the cytokine cluster on chromosome 5q31 (Rioux, Daly et al. 2001). The GWASs carried out in recent years have been considerably more successful, allowing the identification of more than 65 novel loci (Table 1.6.3.2.1). It is interesting to note that, similarly to what has been observed in psoriasis, many of the disease associated markers lie in genes that are related to IL-23/IL-17 and NF-κB signalling (Table 1.6.3.2.1). T cell activation and autophagy are among the other

pathways implicated by GWASs (Table 1.6.3.2.1). The latter observation is of considerable interest, given the well-document role of autophagy in innate and adaptive immune responses (Levine and Deretic 2007).

Table 1.6.3.2.1: Loci associated with CD susceptibility

Gene*	Chr.	Protein function**	Cellular process	References
PTPN22	1p13	Protein tyrosine phosphatase	T cell activation	(The Wellcome Trust Case-Control Consortium 2007; Franke, McGovern et al. 2010)
IL23R	1p31	Subunit of the IL-23 receptor	IL-23 signalling	(Duerr, Taylor et al. 2006; Franke, Hampe et al. 2007; Libioulle, Louis et al. 2007; Parkes, Barrett et al. 2007; Raelson, Little et al. 2007; Rioux, Xavier et al. 2007; The Wellcome Trust Case-Control Consortium 2007; Barrett, Hansoul et al. 2008; Franke, McGovern et al. 2010; McGovern, Jones et al. 2010)
VAMP3	1p36	Member of the vesicle-associated membrane protein (VAMP)/synaptobrevin family	Protein trafficking	(Franke, McGovern et al. 2010)
MUC1	1q22	Mucine	Epithelial cell protection	(Franke, McGovern et al. 2010)
ITLN1	1q23	Lectin	Glucose transport	(Franke, McGovern et al. 2010)
None	1q24	-	-	(Barrett, Hansoul et al. 2008; Franke, McGovern et al. 2010)

DENND1B	1q31	Guanine nucleotide exchange factors	Endocytosis	(Franke, McGovern et al. 2010; McGovern, Jones et al. 2010)
KIF21B	1q32	Microtubule-binding protein	Unknown	(Franke, McGovern et al. 2010; McGovern, Jones et al. 2010)
IL10	1q32	Inhibitory cytokine	NF-κB and JAK/STAT signalling	(Franke, McGovern et al. 2010)
REL	2p16	Subunit of the NF-κB transcription factor	NF-κB signalling	(Franke, McGovern et al. 2010)
THADA	2p21	Unknown	Unknown	(Franke, McGovern et al. 2010)
DNMT3A	2p23	DNA methyltransferase	Methylation establishment	(Franke, McGovern et al. 2010)
GCKR	2p23	Glucokinase regulator	Glucose transport	(Franke, McGovern et al. 2010)
IL18RAP	2q12	Accessory subunit of the IL-18 receptor	NF-κB and JNK signalling	(Franke, McGovern et al. 2010; McGovern, Jones et al. 2010)
PLCL1	2q33	Phospholipase	Inositol phospholipid cascade	(Franke, McGovern et al. 2010)
ATG16L1	2q37	Part of a large protein complex required for the degradation of intracellular components	Autophagy	(Burton, Clayton et al. 2007; Franke, Hampe et al. 2007; Hampe, Franke et al. 2007; Libioulle, Louis et al. 2007; Parkes, Barrett et al. 2007; Raelson, Little et al. 2007; Rioux, Xavier et al.

				2007; Barrett, Hansoul et al. 2008; Franke, McGovern et al. 2010; McGovern, Jones et al. 2010)
SP140	2q37	Nuclear body protein	Unknown	(Franke, McGovern et al. 2010)
MST1	3p21	Unknown	Unknown	(Barrett, Hansoul et al. 2008; Franke, McGovern et al. 2010)
PTGER4	5p13	Member of G protein receptor family	T cell activation	(Franke, Hampe et al. 2007; Libioulle, Louis et al. 2007; Raelson, Little et al. 2007; Rioux, Xavier et al. 2007; The Wellcome Trust Case-Control Consortium 2007; Barrett, Hansoul et al. 2008; Franke, McGovern et al. 2010; McGovern, Jones et al. 2010)
ERAP2	5q15	Amino peptidase processing MHC class I ligands	Antigen presentation	(Franke, McGovern et al. 2010)
SLC22A4	5q31	Organic cation transporter	Unknown	(Franke, Hampe et al. 2007; Hampe, Franke et al. 2007; Barrett, Hansoul et al. 2008; Franke, McGovern et al. 2010)

NDFIP1	5q31	Nedd-4 family interacting protein	Protein trafficking	(Franke, McGovern et al. 2010)
IRGM	5q33	Putative GTPase	Autophagy	(Parkes, Barrett et al. 2007; Raelson, Little et al. 2007; The Wellcome Trust Case-Control Consortium 2007; Barrett, Hansoul et al. 2008; Franke, McGovern et al. 2010)
IL12B	5q33	Subunit shared by the IL-12 and IL-23 cytokines	IL-23 signalling	(Parkes, Barrett et al. 2007; Barrett, Hansoul et al. 2008; Franke, McGovern et al. 2010; McGovern, Jones et al. 2010)
CPEB4	5q35	Unknown	Unknown	(Franke, McGovern et al. 2010)
HLA-DQA2	6p21	MHC class II antigen	Antigen presentation	(van Heel, Fisher et al. 2004; Franke, McGovern et al. 2010)
CDKAL1	6p22	Unknown	Unknown	(The Wellcome Trust Case-Control Consortium 2007; Barrett, Hansoul et al. 2008; Franke, McGovern et al. 2010; McGovern, Jones et al. 2010)
BACH2	6q15	Transcription factor	MAFK signalling	(Franke, McGovern et al. 2010)

PRDM1	6q21	Inhibitor of beta-interferon gene expression	B cell maturation	(Franke, McGovern et al. 2010)
TAGAP	6q25	Putative Rho GTPase-activating protein	T cell activation	(Franke, McGovern et al. 2010)
CCR6	6q27	Chemokine receptor	B cell maturation	(Barrett, Hansoul et al. 2008; Franke, McGovern et al. 2010; McGovern, Jones et al. 2010)
-	7p12	-	-	(Barrett, Hansoul et al. 2008; Franke, McGovern et al. 2010)
JAK2	9p24	Protein tyrosine kinase	IL-23 and IFN mediated JAK/STAT signalling	(Barrett, Hansoul et al. 2008; Franke, McGovern et al. 2010; McGovern, Jones et al. 2010)
TNFSF15	9q32	Cytokine	NF-κB and MAPK signalling	(Yamazaki, McGovern et al. 2005; The Wellcome Trust Case-Control Consortium 2007; Barrett, Hansoul et al. 2008; Franke, McGovern et al. 2010)
CARD9	9q34	Activates NF-κB *via* BCL10	NF-κB signalling	(Franke, McGovern et al. 2010; McGovern, Jones et al. 2010)
CREM	10p11	Transcriptional regulator binding the cAMP response element (CRE)	cAMP signalling	(Franke, McGovern et al. 2010)

IL2RA	10p15	Subunit of the IL-2 receptor	IL-2 signalling	(The Wellcome Trust Case-Control Consortium 2007; Franke, McGovern et al. 2010)
ZNF365	10q21	Unknown	Unknown	(Rioux, Xavier et al. 2007; The Wellcome Trust Case-Control Consortium 2007; Barrett, Hansoul et al. 2008; Franke, McGovern et al. 2010)
UBE2D1	10q21	Ubiquiting-conjugating enzyme	Proteosomal degradation	(Franke, McGovern et al. 2010)
ZMIZ1	10q22	Regulator of transcription factor activity	Androgen receptor signalling	(Franke, McGovern et al. 2010)
NKX2-3	10q24	Member of the NKX family of homeodomain-containing transcription factors	Unknown	(Parkes, Barrett et al. 2007; The Wellcome Trust Case-Control Consortium 2007; Barrett, Hansoul et al. 2008; Franke, McGovern et al. 2010; McGovern, Jones et al. 2010)
FADS1	11q12	Member of the fatty acid desaturase (FADS) gene family	Lipid metabolism	(Franke, McGovern et al. 2010)
C11orf30	11q13	Unknown	Unknown	(Barrett, Hansoul et al. 2008; Franke, McGovern et al. 2010; McGovern, Jones et al. 2010)

PRDX5, ESRRA	11q13	*PRDX5*: member of the peroxiredoxin family of antioxidant enzymes *ESRRA*: nuclear receptor protein	*PRDX5*: intracellular redox signalling *ESRRA*: unknown	(Franke, McGovern et al. 2010)
MUC19	12q12	Mucine	Epithelial cell protection	(Franke, McGovern et al. 2010)
C13orf31	13q14	Unknown	Unknown	(Franke, McGovern et al. 2010)
TNFSF11	13q14	Cytokine	T cell proliferation	(Franke, McGovern et al. 2010)
ZFP36L1	14q24	Unknown	Unknown	(Franke, McGovern et al. 2010)
GALC, GPR65	14q35	*GALC*: galactosylceramidase *GPR65*: glycosphingolipid receptor	*GALC*: lysosomal catabolism *GPR65*: possible involvement in T cell differentiation	(Franke, McGovern et al. 2010)
SMAD3	15q22	Transcription factor	TGF-β signalling	(Franke, McGovern et al. 2010)
IL27	16p11	Cytokine	T cell activation	(Franke, McGovern et al. 2010)
NOD2/CARD15	16q12	NOD-like receptor	NF-κB signalling	(Hugot, Chamaillard et al. 2001; Ogura, Bonen et al. 2001; Franke, Hampe et al. 2007; Hampe, Franke et al. 2007; Libioulle, Louis et al. 2007; Parkes, Barrett et al. 2007; Rioux, Xavier et al. 2007; The Wellcome Trust Case-Control Consortium

				2007; Barrett, Hansoul et al. 2008; Franke, McGovern et al. 2010)
CCL2/CCL7	17q12	Chemokines	JAK/STAT signalling	(Franke, McGovern et al. 2010; McGovern, Jones et al. 2010)
ORMDL3	17q21	Negative regulator of sphingolipid synthesis	Ca^{2+} signalling	(Barrett, Hansoul et al. 2008; Franke, McGovern et al. 2010; McGovern, Jones et al. 2010)
STAT3	17q21	Transcription factor	JAK/STAT signalling	(The Wellcome Trust Case-Control Consortium 2007; Barrett, Hansoul et al. 2008; Franke, McGovern et al. 2010; McGovern, Jones et al. 2010)
PTPN2	18p11	Protein tyrosine phosphatase	IFN signalling	(Parkes, Barrett et al. 2007; The Wellcome Trust Case-Control Consortium 2007; Barrett, Hansoul et al. 2008; Franke, McGovern et al. 2010; McGovern, Jones et al. 2010)

SBNO2	19p13	Transcriptional repressor	Macrophage activation	(Barrett, Hansoul et al. 2008; Franke, McGovern et al. 2010; McGovern, Jones et al. 2010)
TYK2	19p13	Tyrosine kinase associated with cytokine receptors	IL-23 and IFN signalling	(Franke, McGovern et al. 2010)
FUT2	19q13	Fucosyl-transferase	Blood antigen biosynthesis	(Franke, McGovern et al. 2010; McGovern, Jones et al. 2010)
TNFRSF6B	20q13	Member of the tumour necrosis factor receptor superfamily	Apoptosis	(Franke, McGovern et al. 2010)
ICOSLG	21q22	Ligand for the T cell surface receptor ICOS	T cell activation	(Barrett, Hansoul et al. 2008; Franke, McGovern et al. 2010)
YDJC	22q11	Unknown	Unknown	(Franke, McGovern et al. 2010)
MTMR3	22q12	Lipid and protein phosphatase	Autophagy	(Franke, McGovern et al. 2010)
MAP3K7IP1	22q13	MAP kinase regulator	TGF-β signalling	(Franke, McGovern et al. 2010; McGovern, Jones et al. 2010)

*Gene of interest found in the disease-associated susceptibility interval; **Function annotated in the Uniprot protein database.
Adapted from (Franke, McGovern et al. 2010)

1.7 Aim and overview of the project

The hypothesis driving this study is that CD susceptibility genes may contribute to the pathogenesis of psoriasis by affecting shared inflammatory pathways. This notion is supported by the results of early GWASs, which demonstrated the existence of pleiotropic susceptibility genes predisposing to the onset of clinically distinct disorders (The Wellcome Trust Case-Control Consortium 2007). More recently, psoriasis and CD GWASs have identified *IL23R* and *IL12B* as genes conferring susceptibility to both conditions (Duerr, Taylor et al. 2006; Capon, Di Meglio et al. 2007; Cargill, Schrodi et al. 2007; Barrett, Hansoul et al. 2008). Of note, both *IL12B* and *IL23R* are involved in the processes leading to the activation of Th17 lymphocytes (Di Cesare, Di Meglio et al. 2009), a T cell subset that plays a key role in the pathogenesis of chronic inflammation (Martinez, Nurieva et al. 2008).

To address the hypothesis that further CD variants may be associated with psoriasis susceptibility, 26 CD loci identified prior to the onset of this study were examined in a well characterized clinical resource, including 1,256 psoriatic patients and 2,938 unrelated controls. Positive findings were followed-up in an independently ascertained dataset, including 1,348 cases and 1,368 controls. These analyses identified an association between psoriasis and a SNP lying within the *CDKAL1* gene.

To better characterize the function of *CDKAL1*, the gene expression pattern was analysed in multiple tissues and cell types. Stable *CDKAL1* knock-down cell lines were also generated and characterized by a variety of methods, including cell cycle analysis and transcriptome profiling. The results of these experiments indicated that *CDKAL1* may be exerting its pathogenic effect by influencing cell cycle progression and immune responses.

Chapter 2: Materials and methods

2.1 Materials

2.1.1 General reagents

Reagents	**Supplier**
4',6-diamidino-2-phenylindole (DAPI)	Invitrogen
Acetic acid	BDH
BigDye 3.1	Applied Biosystems
Bovine serum albumin (BSA)	GE Healthcare
Coating matrix	Cascade Biologics
Cy-3-Streptavidin	GE Healthcare
Deoxynucleotide Triphosphate (dNTP)	Abgene
Dimethyl sulphoxide (DMSO)	Sigma-Aldrich
Dithiothreitol (DTT)	Invitrogen
Dyethyl pyrocarbonate (DEPC)	Sigma-Aldrich
Ethanol	BDH
Ethylenediamine tetraacetic acid (EDTA)	BDH
ExoSAP-IT	GE Healthcare
Glycerol	BDH
Goat serum	Dako
HiDi formamide	Applied Biosystems
Human Blood Fractions MTC cDNA Panel	Clontech
Human MTC cDNA Panel I	Clontech
Human MTC cDNA Panel II	Clontech
Hydrogen peroxide (H_2O_2)	BDH

Image clone 30379056	Source BioScience
Isopropanol	BDH
Isopropyl B-D-1-thiogalactopyranoside (IPTG)	Sigma-Aldrich
Lipofectamine 2000	Invitrogen
Magnesium chloride ($MgCl_2$)	BDH
Methanol	Sigma-Aldrich
N-2-hyroxyethylpiperazine-N′-2-ethanesulphonic acid (HEPES)	Sigma-Aldrich
Nonidep-P40	Sigma-Aldrich
ON-TARGETplus SMART pool siRNA, CDKAL1	Thermo Scientific
Paraformaldehyde	Sigma-Aldrich
pGEX-4T-1	GE Healthcare
Phenol: chloroform: isoamyl alcohol	Sigma-Aldrich
Polyethilenimine (PEI)	Sigma-Aldrich
Ponceau S	Sigma-Aldrich
Potassium choride (KCl)	BDH
Prolong Gold AntiFade Reagent	Invitrogen
Propidium iodide (PI)	BDH
Protease inhibitor cocktail tablets	Sigma-Aldrich
Puromycin	Sigma-Aldrich
Pytohemagglutinin-M (PHA)	Roche
Sequabrene	Sigma-Aldrich
sh-non-silencing	Open Biosystems
sh-RNA-V2LHS_17373 individual clone	Open Biosystems
sh-RNA-V2LHS_173734 individual clone	Open Biosystems

Skimmed milk	Morrison
Sodium acetate (NaOAc)	BDH
Sodium chloride (NaCl)	Sigma-Aldrich
Sodium dodecyl sulphate (SDS)	Sigma-Aldrich
Sodium hydroxide (NaOH)	BDH
TaqMan Gene expression assays	Applied Biosystems
TaqMan Genotyping master mix	Applied Biosystems
TaqMan SNP Genotyping assays	Applied Biosystems
Tetrazolium salt (WST-8)	Roche
Triton X-100	Sigma-Aldrich
Tween 20	Sigma-Aldrich
Vaseline	Morrison

2.1.2 General enzymes and related buffers

Reagent	**Supplier**
10X ligase buffer	New England Biolabs
10X NEBuffer 4	New England Biolabs
10X PCR buffer	New England Biolabs
Calf Intestinal alkaline phosphatase (CIP)	New England Biolabs
Dispase	StemCell Technologies Inc
EcoRI-HF	New England Biolabs
Proteinase K	Sigma-Aldrich
Ribonuclease A (RNase)	Sigma-Aldrich
T4 DNA ligase	New England Biolabs
Taq polymerase	Sigma-Aldrich
Trypsin-EDTA	Invitrogen
XhoI	New England Biolabs

2.1.3 Electrophoresis gel and loading buffer constituents

Reagent	**Supplier**
5 X DNA loading buffer	Bioline
Agarose	BDH
Ammonium persulphate (APS)	Sigma-Aldrich
Ethidium bromide	Sigma-Aldrich
Hyperladder I molecular weight marker	Bioline
Protein Loading buffer	Sigma-Aldrich
Sizing ladder (precision Plus Protein Standards)	BIORAD
Tetramethylethylenediamine (TEMED)	Sigma-Aldrich
Ultra Pure ProtoGel 30% acrylamide mix	National Diagnostics

2.1.4 Bacterial and tissue culture growth medium constituents

Reagent	**Supplier**
Ampicillin	Sigma-Aldrich
Bacto-Agar	BD
Bacto-tryptone	BD
Dulbecco's Modified Eagle's medium (DMEM)	Invitrogen
Epilife keratinocyte medium	Cascade biologics
Foetal Calf Serum (FCS)	Invitrogen
Human Keratinocyte Growth Supplement	Cascade biologics
L-glutamine	Invitrogen
Lipofectamine 2000	Invitogen
Penicillin-Streptomycin (P/S)	Invitrogen
Phosphate buffer saline (PBS)	Invitrogen
Roswell Park Memorial Institute (RPMI) 1640	Invitrogen
Trypsin-EDTA	Invitrogen
Yeast extract	BD

2.1.5 Cytokines

Reagent	**Supplier**
Interferon-alpha (IFN-α)	R&D Systems
Interferon-gamma (IFN-γ)	R&D Systems
Interleukin-17 (IL-17)	R&D Systems
Interleukin-2 (IL-2)	R&D Systems
Interleukin-22 (IL-22)	R&D Systems
Interleukin-23 (IL-23)	R&D Systems
Tumor necrosis factor-alpha (TNF-α)	R&D Systems

2.1.6 Antibodies

Reagent	**Supplier**
Alexa Fluor 555 goat anti-mouse IgG (H+L) F (Ab) 2 fragments	Invitrogen
Donkey Anti-Goat IgG HRP	Santa Cruz
Goat Anti-Mouse IgG_HRP	Dako
Goat Anti-Rabbit IgG _HRP	New England Biolabs
α- CDKAL1	Sigma-Aldrich
α- Lamin alpha	Sigma-Aldrich
α- β actin	New England Biolabs

2.1.7 Molecular biology kits

Reagent	**Supplier**
Bicinchoninic Acid Protein Assay (BCA) Kit	Sigma-Aldrich
Enhanced chemiluminescence (ECL)	GE Healthcare
Gene Expression Direct Hybridization Assay Kit	Illumina
HumanHT-12_v3_BeadChip	Illumina
QIAprep Maxiprep Kit	Qiagen
QIAprep Miniprep Kit	Qiagen
QIAquick PCR purification Kit	Qiagen
Quanti-iT RNA assay Kit	Invitrogen
QuantiTect Rev. Transcription Kit	Qiagen
RNA Amplification kit	Illumina
RNeasy Plus Mini Kit	Qiagen
TotalPrep RNA Kit	Illumina
TRI REAGENT Kit	Sigma-Aldrich

2.1.8 Stock solutions

Solution	**Constituents**
1X Tris-EDTA buffer (TE)	10 mM Tris-HCl, pH 8.0 1mM EDTA
10X Running buffer	250 mM Tris base 1.92 M Glycine 1% (w/v) SDS
10X Tris-Borate-EDTA buffer (TBE)	90 mM Tris base 90 mM Boric acid 1 mM EDTA
Coomassie blue stain solution	0.5% (w/v) Coomassie blue 40% (v/v) Methanol
Chorionic Villus Sample buffer (CVS)	10 mM Tris base 1 mM EDTA 0.1 M NaCl 0.5% (v/v) SDS Adjust pH to 10.5 with NaOH
Cytoplasmic cell lysis buffer	1mM PMSF 10 mM HEPES pH 7.9 10 mM KCl 0.1 mM EDTA 0.1 mM EGTA 1 mM DTT

De-staining solution	40% (v/v) MeOH 10% (v/v) NaOAc
Luria broth (LB)/agar (LUA)	1% (w/v) Bacto-Tryptone 1% (w/v) NaCl 0.5% (w/v) Yeast extract For LUA, add 1.5 % (w/v) Bacto-Agar
Nuclear cell lysis buffer	1mM PMSF 20 mM HEPES pH 7.9 0.4 M NaCl 1 mM EDTA 1 mM EGTA 1 mM DTT
Ponceau S	0.1% (w/v) Ponceau S 5% (v/v) Glacial acetic acid
Sequencing precipitation solution	95% EtOH 120 mM NaOAc, pH 4.6
Sucrose lysis buffer	0.3 M Sucrose 10 mM Tris-HCl, pH 7.5 5 mM $MgCl_2$ 1% Triton X-100
Total cell extract buffer	100 mM Tris-HCl, pH 6.8 4% SDS

Transfer buffer	25 mM Tris base 192 mM Glycine 20% (v/v) MeOH (add immediately before use)
Transformation buffer (TB)	15 mM $CaCl_2$ 250 mM KCl 10 mM PIPES 55 mM $MnCl_2$
10X Tris Buffered Saline (TBS)	1.5 M NaCl 0.2 M Tris base, pH 7.6

2.2 Subjects

2.2.1 Study subjects

A total of 1,256 British patients of North European descent were enrolled in this study. All cases were ascertained from specialist dermatology clinics, at St John's Institute of Dermatology, London (n= 638), Glasgow Western Infirmary (n= 211), and the Dermatology Centre, University of Manchester (n= 407). Experienced dermatologists from each institution confirmed that all patients showed early onset (occurring before 40 years of age) psoriasis vulgaris, defined according to established clinical criteria (Camp 1998). Less than 1% of the patients were also known to suffer from CD.

The control individuals (n= 2,938) had been previously analysed by the Wellcome Trust Case-Control Consortium (WTCCC) and originated from two sources, the 1958 British Birth Cohort and the UK Blood Service (The Wellcome Trust Case-Control Consortium 2007) .

A total of 1,348 cases and 1,368 unrelated controls from the Collaborative Association Study of Psoriasis (CASP) were used as a replication dataset. Both cases and controls were recruited in the US among individuals of North-European descent (Nair, Duffin et al. 2009). Human studies were conducted in accordance with the Helsinki Declaration and approved by the Ethics Committees of all participating institutions. All cases and controls gave their informed consent for the use of their DNA in genetic epidemiology analyses.

2.2.2 DNA extraction from peripheral blood

Genomic DNA was extracted from peripheral blood leukocytes using the "salting out" method (Miller, Dykes et al. 1988). Ten milliliters of blood were collected in EDTA containers, dispensed in 50 mL polypropylene tubes and mixed with 40 mL of ice-cold

distilled water (dH_2O). Samples were centrifuged at 3,000 rpm for 10 min at 4˚C. The supernatant was discarded and the pellet re-suspended in 25 mL of ice-cold sucrose lysis buffer. Samples were again centrifuged at 3,000 rpm for 10 min at 4˚C. The supernatant was discarded and the pellet re-suspended in 3 mL CVS buffer supplemented with 25 µL of 20 mg/mL proteinase K. Samples were incubated at 60˚C for 2-3 hours or overnight at 37˚C and then vortexed to ensure that the pellet was completely dissolved. Subsequently, 1.8 mL of 5 M NaCl was added to the solution and tubes were centrifuged at 3,500 rpm for 30 min at 4˚C. The supernatant was transferred to clean 15 mL polypropylene tubes and DNA was precipitated by adding 5 mL 90% ethanol. DNA was transferred into 1.5 mL tubes containing 1 mL 70% ethanol and centrifuged at 3,000 rpm for 10 min. After discarding the supernatant the pellet was resuspended in 1 mL of TE pH8.0.

2.2.3 DNA plating

DNA samples were quantified by measurement of their OD_{260} with a NanoDrop ND1000 Spectrophotometer (Labtech). DNAs were aliquoted into 96 well plates at a concentration of 25 ng/µL. Positive (two duplicate samples in each plate) and negative controls were also plated, in order to verify the reproducibility of genotyping results and to detect the presence of cross-contamination.

2.3 DNA typing

2.3.1 Single Nucleotide Polymorphism genotyping

Patient DNAs were typed using TaqMan SNP Genotyping Assays. These reagents allow accurate genotypic discrimination based on the use of allele-specific fluorochromes (VIC for allele 1 and FAM for allele 2) linked to the 5' end of two oligonucleotide

probes. Each TaqMan probe also contains a minor groove binder (MGB) at the 3' end (allowing the user to increase the melting temperature and to design shorter probes) and a quencher that acts by decreasing background fluorescence. Allele specific signals are detected when the probes are incorporated into the PCR product and the quencher is cleaved.

All genotyping reactions were performed in 96 well plates covered by optical adhesive lids. A master mix was prepared to a final volume of 260 µL, including 130 µL 2X TaqMan Genotyping master mix, 6.5 µL TaqMan Genotyping Assay and 123.5 µL dH_2O. 2.5 µL of master mix were dispensed in each well and 1 µL of DNA (25 ng/µL) was added. PCR reactions were carried out in a BIORAD PTC-100 Peltier thermal cycler, under the following conditions: 15 min 95°C; (15 sec 92°C, 1 min 60°C) x 44. Amplification products were visualized on an Applied Biosystems Real-time PCR 7900HT apparatus, using the SDS 2.2 software .

2.3.2 Direct sequencing

2.3.2.1 Primer design

Primers amplifying the regions of interest (Table 2.3.2.1.1) were designed using the publicly accessible Primer3 software (http://frodo.wi.mit.edu/primer3/).

Table 2.3.2.1.1: Primer sequences

Target	**Sequences**
pGEX-4T-1-CDKAL1	5'-CGCGAATTCAATATGCCTTCTGCATCC-3` 5'- GCGCTCGAGAGTACAAATTATTGCCCT-3`
CDKAL1 full-length cDNA	5'-TGCAGATTTATGGCTCCTGA-3` 5'- AGCAAAAAGAAGACCCAGCA-3`
Region surrounding SNP rs11175593	5'-TGGTCAGCACTCTTGCACAT-3` 5'-AAGGCAATATGCTTTAGGATTCA-3`

2.3.2.2 Polymerase Chain Reaction (PCR)

PCR reactions were carried out in a 20 µL final volume in the presence of 0.25 µM forward and reverse oligonucleotide primers, 1X Taq PCR reaction buffer (75 mM Tris-HCl (pH8.8), 20 mM NH_4SO_4, 1.5 mM $MgCl_2$, 0.01% (v/v) Tween-20), 0.2 mM dNTPs, and 0.05 U/µL Taq polymerase. The amplification reaction was carried out in a BIORAD PTC-100 thermal cycler under the following conditions: 96˚C for 2 min (96°C 30 sec; 55°C 30 sec; 72°C 30 sec) x 30; 72°C 5 min. Three microlitres of PCR product plus 3 µL of 5X loading dye were loaded on a 2% agarose gel containing 0.5 µg/mL ethidium bromide, alongside the HyperLadder I DNA sizing marker. Gels were electrophoresed at 130 V in 1X TBE buffer and PCR products were visualized on an UV transilluminator.

2.3.2.3 Direct sequencing

Sequencing reactions were performed in 96 well plates. PCR products were purified using the ExoSAP-IT reagent, a mixture of exonuclease and shrimp alkaline phosphatase that removes the excess of primers and nucleotides used in the amplification reaction. Two microlitres of PCR product were mixed with 3 µL of ExoSAP-IT and incubated at 37°C for 30 min, followed by an additional 15 min at 80°C. Amplicons were sequenced with 2 µM primer (Table 2.3.2.1.1), using 1.25 µL of 5X buffer and 0.25 µL BigDye 3.1 reagent. After 30 sequencing cycles (96°C 30 sec; 50°C 15 sec; 60°C 60 sec), reactions were purified by ethanol precipitation. Briefly, 26 µL of precipitation solution was added to each reaction. Samples were incubated at room temperature for 10 min and centrifuged at 3,000 rpm for 30 min. The supernatant was removed by a brief centrifugation of the inverted plate at 400 rpm. Samples were washed with 70 µL of 70% ethanol and further centrifuged at 3000 rpm for 10 min.

Ethanol was removed and the pellet was air-dried for 10 min, re-suspended in 10 µL of HiDi formamide and denatured at 90˚C for 2 min. Samples were loaded on an ABI 3730*xl* sequencer (Applied Biosystems). Sequence analysis was conducted by visual inspection of chromatograms, using both the FinchTV (available at www.geospiza.com/finchtv.html) and the Sequencher 4.7 (http://www.softpedia.com/get/Science-AD/Sequencher.shtmlprograms) software.

2.4 Cell culture

2.4.1 Culture conditions

2.4.1.1 Adherent immortalized cell lines

All tissue culture procedures were performed under sterile conditions using a laminar flow hood (NUAIRE, Biological Safety Cabinet). *In-vitro* spontaneously transformed keratinocytes from histologically normal skin (HaCaT) and Human Embryonic Kidney 293 cells (HEK293T) cells were grown in a humidified atmosphere at 37˚C in 5% CO_2. Both cell lines were cultured in Dulbecco's Modified Eagle's medium (DMEM) supplemented with 10% Fetal Calf Serum (FCS), 2 mM L-glutamine, 100 U/mL penicillin and 100 µg/mL streptomycin (complete medium). For cryopreservation, aliquots of 10^6 cells were prepared in a final volume of 1 mL 90% FCS/10% Dimethyl Sulphoxide (DMSO) and transferred to liquid nitrogen, after a progressive temperature shift.

Cells were passaged upon reaching 80% confluence. The medium was removed and the cells washed in Phosphate Buffered Saline (PBS) in order to remove any trace of serum. Cells were incubated at 37˚C for 5 min, in 2.5 mL of pre-warmed trypsin-EDTA. The trypsin was inactivated by adding 5 mL fresh complete medium into the flask. Cells were centrifuged at room temperature for 5 min at 1,200 rpm and the pellet resuspended

in 10 mL of fresh medium.

2.4.1.2 Cell lines grown in suspension

Jurkat immortalized T cells were cultured in Roswell Park Memorial Institute (RPMI) 1640 medium supplemented with 10% FCS, 2 mM L-glutamine, 100 U/mL penicillin and 100 µg/mL streptomycin (complete medium). Cells were cryopreserved for storage as described above and were maintained at a concentration of 3-9 x 10^4 cells/mL. When this density was exceeded, cells were centrifuged at room temperature at 1,200 rpm for 5 min. The pellet was washed in PBS and resuspended in 10 mL fresh medium.

2.4.1.3 Primary keratinocytes

Primary keratinocytes were derived from the skin biopsies of three consenting donors (two unaffected individuals and a psoriatic patient). The dermis was cleaned of fat using a scalpel; the skin was cut in small stripes, transferred in a 6 well plate and incubated overnight on a shaker with 5 mg/mL dispase. The epidermis was separated from the dermis with sterile forceps and washed in 4 mL of sterile PBS. The epidermis was then incubated with 4 mL trypsin-EDTA at 37 ˚C for 15 min. The tissue was homogenised using a 5 mL syringe and immediately passed through a 100 µm cell mesh (BD Bioscience) pre-treated with FCS. The cell suspension was transferred to a 10 mL tube and washed in PBS twice. The pellet was re-suspended in 10 mL Epilife keratinocyte medium supplemented with Human Keratinocyte Growth Supplement and transferred to 75 cm^2 flasks pre-coated with coating matrix. Keratinocytes were maintained in culture until a confluent monolayer was formed. Cells were trypsinised as described above and cytokine treatment was carried out between passages 2-4. Cells were quiesced by culturing them in supplement-free medium for 24 hours. Based on published data on the

kinetics of cytokine induced gene expression (Hundhausen, Bertoni et al. 2011), keratinocytes were incubated with 100 ng/mL of the relevant cytokine (either IL-17, IL-22, IL-23, IFN-α, IFN-γ or TNF-α) for 24 hours (5 hours for IFN-α). All treatments were carried out in duplicate.

2.4.2 Immunofluorescence staining

HaCaT cells were cultured on coverslips in a 20 mm culture dish. After reaching 40% confluence the cells were washed twice in PBS and then fixed in 4% paraformaldehyde (pH>7) for 5 min. After an additional PBS wash, the cell membranes were permeabilized in 0.6% H_2O_2, 0.1% Triton X-100 for 10 min at room temperature. Fixed cells were blocked for 20 min at room temperature in a PBS buffer containing 0.1% BSA, 10% goat serum, and 0.1% Tween20. Cells were incubated with a primary α-CDKAL1 antibody, at a final concentration of 1µg/mL for 1.5 hours at room temperature. Following two washes in PBS, a secondary antibody (Alexa Fluor 555 goat anti-mouse IgG (H+L) F (Ab) 2 fragments) was added at a dilution of 1:200 in blocking buffer. After 30 min incubation at room temperature, coverslips were mounted onto optical slides with Prolong Gold AntiFade Reagent supplemented with 4',6-diamidino-2-phenylindole (DAPI). Images were acquired using a fluorescence microscope (Zeiss Axiophot) and analysed with the NIS-Elements BR 3.0 software.

2.5 Cell transfection and transduction

2.5.1 Transient *CDKAL1* RNA interference

HaCaT cells from a confluent culture were counted under an optical microscope (ZEISS) using a haemocytometer. Thirty-five thousand cells were plated into each well of a 24 well plate and grown overnight in complete DMEM. A 33 nM pool of four

RNAi oligonucleotides (ON-TARGETplus SMARTpool siRNA), was incubated in 50 µL serum-free DMEM for 5 min at room temperature. The diluted oligonucleotides were combined with 1 µL Lipofectamine 2000 and incubated at room temperature for an additional 20 min. The oligonucelotide-Lipofectamine complex was added to the medium of each well and mixed by gentle swirling. The medium was replaced with complete DMEM after 24 hours and cells were grown for a further 48-72 hours.

2.5.2 Stable *CDKAL1* RNA interference

2.5.2.1 Lentivirus production

All lentivirus work was performed in level 2 containment facilities. HEK293T cells, grown in 100 mm cell culture dishes, were transfected with 6 µg HIV-gag/pol expressing vector, 1.5 µg VSV-G expressing vector (both obtained from the Addgene plasmid repository, accessible at http://www.addgene.org/pgvec1), 6 µg of the relevant sh-RNA construct (either sh-non-silencing, sh-RNA-V2LHS_173734 individual clone or sh-RNA-V2LHS_17373 individual clone). Plasmids were diluted in 750 µL serum free medium and incubated at room temperature for 5 min. Sixty microlitres of 1 mg/mL polyethylenimine (PEI) were added and the mixture was incubated for an additional 20 min at room temperature. The complex was added to the cell medium and mixed by gentle swirling. After 24 hours the medium was replaced with 8 mL complete DMEM. After a further 48-72 hours the medium was collected, centrifuged at 1,500 rpm for 5 min at room temperature and stored at -80˚C.

2.5.2.2 Cell transduction

Four millilitres of supernatant from transfected HEK293T cells was diluted in 4 mL of complete medium supplemented with 8µg/mL Sequa-brene, and mixed with 2 x 10^6

HaCaT or 4 x 10^6 Jurkat cells. The efficiency of transduction was assessed after 24 hours, by using fluorescence microscopy (see below) to identify the cells expressing tGFP. After a further 24 hours, cells were placed under antibiotic selection, by supplementing the medium with 15 μg/mL puromycin. The selection medium was replaced every 72 hours until puromycin resistant colonies appeared.

2.5.2.3 HaCaT cell clonal expansion

One hundred transduced HaCaT cells were cultured in 20 mm cell culture dishes for 1 week. Well defined clones arising from individual cells were then highlighted with a marker under the microscope. The medium was removed and cells washed twice in PBS. A sterile cloning ring was placed around the marked colonies, using sterile Vaseline to seal it. Two hundred microlitres of pre-warmed trypsin-EDTA was added to the ring and cells were incubated at 37˚C for 5 min. The trypsin-EDTA and cell suspension was added to 5 mL of complete DMEM and centrifuged at 1,200 rpm at room temperature for 5 min. The cell pellet was resuspended in 500 μL compete DMEM and cells were plated in 6 well plates.

2.5.2.4 Jurkat T cell clonal expansion

The clonal expansion of transduced Jurkat T cells was performed in 96 well plates by growing a limiting cell dilution (0.5 cells per well) in the presence of previously irradiated at 50 Gy untransduced Jurkat cells as feeders. Cells were incubated at 37˚C in 5% CO_2 in RPMI 1640 medium supplemented with 2.5 μg/mL Pytohemagglutinin-M (PHA). After a week the medium was replaced with complete RPMI supplemented with 20 IU/mL interleukin-2 (IL-2). Cells were cultured until the appearance of clusters indicated they had reached an appropriate density for expansion.

2.6 Cell cycle analysis

2.6.1 WST-8 proliferation assay

This colorimetric cell viability assay is based on the cleavage of the WST-8 tetrazolium salt into formazan. As this reaction is catalyzed by a mitochondrial dehydrogenase that is more active during cell proliferation, the accumulation of formazan directly correlates with the number of metabolically active cells.

Thirty thousand transduced HaCaT cells, were seeded in triplicate in a 24 well plate and grown in 500 µL complete DMEM. After 24 hours the medium was replaced with 500 µL reduced (1.5%) FCS DMEM and cells were grown for a further 24 hours. The medium was replaced with 500 µL complete DMEM and cells incubated for an additional 24 hours, after which 50 µL of WST-8 were carefully added. The accumulation of formazan was monitored by measuring its absorbance (450 nm) every 30 min on a Fluorostar Omega plate reader (Labtech).

2.6.2 Propidium iodide staining

Propidium iodide (PI) is a fluorescent DNA intercalating agent that is used to identify the proportion of cells in that are in the G_0/ G_1, S or G_2/M phase of the cell cycle.

One million transduced HaCaT cells were washed twice in 1 mL PBS. The pellet was fixed in 1 mL ice-cold 70% ethanol and incubated at 4 °C overnight. Fixed cells were washed twice in PBS and the pellet treated with 50 µg RNase and 20 µg PI, in a final volume of 500 µL. Cells were incubated at room temperature in the dark for 30 min and subsequently analyzed by flow cytometry, using a FACScanto apparatus (BD). Data were analysed with the Flow JO (TreeStar Inc) software.

2.7 Plasmid DNA cloning

2.7.1 Preparation of DH5α competent cells

E.coli DH5α cells were streaked on a Luria broth agar plate (LUA) and incubated at 37°C overnight. One representative colony was inoculated in 2 mL LB medium and grown at 37°C overnight shaking at 160 rpm. One millilitre of culture was added to 500 mL LB medium supplemented with 2 M $MgCl_2$ and incubated at 18°C, until OD600 reached 0.3. The culture was rapidly cooled on ice and cells were centrifuged at 3,000 rpm for 10 min, at 4°C. The cell pellet was resuspended in 150 mL ice cold Transformation Buffer (TB). Cells were further centrifuged at 3,000 rpm for 10 min at 4°C and the pellet was resuspended in 40 mL cold TB. Three millilitres of DMSO were added with gentle swirling. Five-hundred microlitres of cell suspension was aliquoted into 1.5 mL tubes and immediately chilled in liquid nitrogen. Cells were stored at -80 °C.

2.7.2 Digestion

Following an analysis of *CDKAL1* full-length cDNA sequence (accession n.: NM_017774) with NEB cutter v2.0 (http://tools.neb.com/NEBcutter2/index.php), EcoRI and XhoI were selected as suitable enzymes for the cloning procedure. Accordingly, EcoRI- and XhoI- tagged PCR primers were designed (Table 2.3.2.1.1). The full length *CDKAL1* transcript was amplified from Image clone 30379056, using a 10:1 mix of Taq/Pfu polymerase (all other reagents were as described in section 2.3.2.2). An initial denaturation step (96˚C for 2 min) was followed by 30 amplification cycles (96°C 30 sec; 55°C 30 sec; 72°C 1 min) and a final extension (72°C 5 min). The PCR product was column-purified using the QIAquick PCR purification kit according to the manufacturer's protocol. Digestion of PCR product and pGEX-4T-1 vector was carried

out in a final volume of 20 μL, as detailed in table 2.7.2.1. The reaction was incubated at 37˚C for 2 hours. After 1.5 hours 10 IU of Calf Intestinal alkaline Phosphatase (CIP) were added to the vector, in order to prevent its re-circularisation. Digestion products were quantified by visualisation on an agarose gel alongside a Hyperladder I molecular weight marker (see section 2.3.2.2).

Table 2.7.2.1: DNA restriction

Components	**Volume (insert)**	**Volume (vector)**
10X NEBuffer 4	2 µL	2 µL
10X BSA	2 µL	2 µL
EcoRI-HF (20,000U/mL)	0.5 µL	0.5 µL
XhoI (20,000U/mL)	0.5 µL	0.5 µL
Amplification product	6 µL (800 ng)	-
pGEX-4T-1	-	4 µL (1 µg)
H_2O	9 µL	11 µL
Final Volume	20 µL	20 µL

2.7.3 DNA ligation and bacterial transformation

Fifty nanograms of digested pGEX-4T-1 vector were combined with 72 ng of *CDKAL1* RT-PCR product (3:1 insert: vector molar ratio), in a 20 µL ligation reaction containing 2 µL 10X ligase buffer and 1U of T4 DNA ligase. Samples were incubated at 15˚C overnight.

Ligation products were transformed into competent *E.coli* DH5α cells as follows. One-hundred microlitres *E.coli* DH5α cells were slowly defrosted and incubated with the ligation product for 30 min on ice. Cells were heat-shocked at 42˚C, for 90 sec and cooled on ice for a further 2 min. Five-hundred microlitres of LB medium were added to the suspension, and incubated for 1 hour at 37˚C. Samples were centrifuged at 9,000 rpm, for 5 min. The bacterial pellet was resuspended in 100 µL of LB medium and plated onto selective LUA plates containing 50 µg/mL ampicillin. After an overnight incubation at 37˚C, individual colonies were inoculated in 5 mL LB medium supplemented with 50 µg/mL ampicillin. Cultures were incubated at 37°C overnight, with shaking at 160 rpm. Glycerol stocks were prepared by mixing 750 µL of overnight cultures with 250 µL 80% glycerol.

2.7.4 Colony screening

The overnight cultures described above were screened by PCR, in order to confirm the presence of the desired insert. All reactions were carried out in a final volume of 50 µL. Briefly, 2 µL of overnight culture were incubated with 1 µL of Tween 20 and 34.85 µL of dH_2O for 10 min at 95˚C. A master mix containing 5 µL of 1X Taq PCR reaction buffer, 5 µL of 0.2 mM dNTPs, 1 µL of 0.2 µM *CDKAL1* forward and *CDKAL1* reverse primers and 0.15 IU/µL Taq polymerase was added to the mixture. Samples were amplified under the conditions described in section 2.7.2. Amplification products were

visualized by agarose gel electrophoresis.

2.7.5 Plasmid DNA extraction

Overnight cultures were centrifuged for 5 min at 4,600 rpm at 4˚C. Plasmid DNA was extracted using QIAprep Miniprep or QIAprep Maxiprep kits according to the manufacturer's protocol. The concentration of plasmid DNA was quantified with a Nanodrop ND-1000 spectrophotometer. Plasmids were sequenced as described above, in order to verify the integrity and orientation of the insert.

2.8 Transcript analysis

2.8.1 RNA extraction

Total RNA was isolated from cultured cells, using either the RNeasy Plus Mini Kit or TRI REAGENT, according to the manufacturer's protocol. The RNA yield was determined by quantifying the samples on a Nanodrop ND1000 UV-vis Spectrophotometer . The concentration of the RNA samples to be used in the microarray experiment was further confirmed by Quanti-iT RNA assay using a Qubit fluorometer (Invitrogen). The quality of these RNAs was also assessed by using a Bioanalyzer 2100 (Agilent) to monitor the presence of degradation products within the sample.

2.8.2 cDNA and cRNA synthesis

cDNA was synthesised from 500 ng total RNA, using the QuantiTect Rev. Transcription Kit. cRNA was transcribed from 500 ng total RNA, using Illumina's TotalPrep RNA Amplification kit. Both kits were used according to the manufacturers' instructions.

2.8.3 Hybridisation of gene expression chips

Samples were hybridised to a HumanHT-12_v3_BeadChip (Illumina), using Illuminas's Whole-Genome Gene Expression Direct Hybridization Assay Kit, according to the manufacturer's instructions.

2.8.4 Qualitative Reverse Transcription (RT)-PCR

Two microlitres of cDNA were amplified in a in a 20 μL final volume (see section 2.3.2.2 for a list of reaction components and Table 2.3.2.1.1 for primer sequences). An initial denaturation step (96˚C for 2 min) was followed by 38 amplification cycles (96°C 30 sec; 55°C 30 sec; 72°C 30 sec) and a final extension (72°C 5 min). Three microlitres of PCR product were loaded on agarose gel and visualized as described in section 2.3.2.2

2.8.5 Quantitative RT-PCR

Gene expression was assessed by singleplex real-time quantitative RT-PCR using Taqman Gene Expression Assays according to the manufacturer's instructions. All PCR reactions were carried out in duplicate in 96 well plates covered by optical adhesive lids. Samples were amplified in an Applied Biosystems 7900HT Sequence Detection System, under the following conditions: 15 min 95°C; (30s 95°C, 1 min 60°C) x 40. The relative *CDKAL1* expression levels were analyzed using the ΔΔCt method as described in section 2.11.3.

2.9 Protein analysis

2.9.1 Total cell lysis of adherent cells

Cells were cultured in 24 well plates in complete DMEM, until they reached 80-90%

confluence. Cells were washed in PBS and treated with 150 µL of ice cold lysis buffer, supplemented with a cocktail of protease inhibitors (Complete, Mini, EDTA-free tablets). The plate was incubated for 20 min at 4˚C with gentle shaking. Next, the cells were scraped and transferred to a micro-centrifuge tube. The lysate was clarified by centrifuging samples for 10 min at 4,600 rpm, at 4˚C. The supernatant was transferred to a new micro-centrifuge tube and stored at -80˚C.

2.9.2 Total cell lysis of suspension cells

Three million cells were centrifuged 1,200 rpm at room temperature, for 5 min. The supernatant was removed and the pellet was washed in 5 mL ice-cold PBS. Following an additional centrifugation, the pellet was resuspended in 300 µL ice cold cell lysis buffer supplemented with protease inhibitors. Cells were incubated for 30 min on ice and briefly vortexed. The supernatant was transferred to a micro-centrifuge tube and centrifuged for 10 min at 4,600 rpm, at 4˚C. The supernatant was transferred to a fresh tube and stored at -80˚C.

2.9.3 Nuclear and cytoplasmic protein extraction

The pellet from 3 x 10^6 adherent or 5 x 10^6 suspension cells was re-suspended in 400 µL ice cold cytoplasmic lysis buffer and left on ice for 15 min. Cells were supplemented with 10% NP40, vortexed vigorously for 20 sec, incubated on ice for 20 min and finally centrifuged at 4,600 rpm for 10 min at 4˚C. The supernatant was collected without disturbing the nuclear pellet, transferred to a new 1.5 mL pre-chilled tube and stored at -80˚C. The nuclear pellet was washed several times in cytoplasmic lysis buffer and centrifuged at 1,600 rpm at 4˚C for 10 min. The nuclear pellet was resuspended in 400 µL ice cold nuclear cell lysis buffer and incubated at 4˚C overnight, in a shaker. The

suspension was centrifuged at 1,600 rpm at 4˚C for 20 min. The supernatant was transferred to a pre-chilled 1.5 mL tube and stored at -80˚C.

2.9.4 Protein quantification

All protein extracts were quantified using the Bicinchoninic Acid Protein Assay (BCA) kit. The assay is based on the reduction of the Cu^{2+}-protein complex into Cu^{1+}, which, under alkaline conditions results in the formation of a purple-blue complex (Smith, Krohn et al. 1985). Samples were treated according to the manufacturer's protocol and the colorimetric reaction was quantified on a Fluorostar Omega plate reader at 562 nm.

2.9.5 Western blotting

Twenty micrograms of protein extract were mixed with gel loading buffer at a 1:1 ratio. Samples were denatured at 90˚C for 5 min, chilled on ice and electrophoresed in a 10% discontinuous SDS-polyacrylamide minigel (see Tables 2.9.5.1 and 2.9.5.2), in the presence of 7 µL protein sizing ladder. Proteins were separated at 120V for 2 hours and subsequently transferred onto a Protran Nitrocellulose membrane, using the Trans-blot SD Semi-dry Transfer cell blotting system (BIORAD), at 250 mA for 40 min. Transfer efficiency was checked by Ponceau red staining of proteins and membranes were blocked for 2 hours in 5% skimmed milk diluted in 0.1% Tween 20_Tris Buffered Saline (TBS). The membranes were incubated overnight at 4˚C with the relevant antibody, washed 4 times in TBS-0.1% Tween 20 and incubated for 2 hours at room temperature with the relevant HRP-conjugated secondary antibody (see Table 2.9.5.3). After a second round of washing, membranes were stained for 5 min at room temperature with ECLplus. Protein bands were detected by autoradiography.

Table 2.9.5.1: **Separating gel components**

Components	**10% gel**	**8% gel**
H_2O	5.9 mL	6.9 mL
Ultra Pure ProtoGel 30% acrylamide mix	5.0 mL	4.0 mL
1.5M Tris (pH8.8)	3.8 mL	3.8 mL
10% SDS	150 µL	150 µL
10% ammonium persulfate	150 µL	150 µL
TEMED	6 µl	6 µL
Final volume	15 mL	15 mL

Table 2.9.5.2: 5% Stacking gel components

Components	**Volumes**
H_2O	2.1 mL
Ultra Pure ProtoGel 30% acrylamide mix	500 µL
1.5M Tris (pH6.8)	380 µL
10% SDS	30 µL
10% ammonium persulfate	30 µL
TEMED	3 µL
Final volume	3 mL

Table 2.9.5.3: **Antibodies used in western blot experiments**

Primary antibodies	**Dilutions**	**Secondary antibodies**	**Dilutions**
α-CDKAL1	1:1,000	Goat Anti-Rabbit IgG HRP	1:2,000
α- β actin	1:10,000	Goat Anti-Mouse IgG/HRP	1:2,500
α- Lamin alpha	1:1,000	Donkey Anti-Goat IgG HRP	1:5,000

2.9.6 Isopropyl-β-D-1-thiogalactopyranoside (IPTG) induction

Isopropyl β-D-1-thiogalactopyranoside (IPTG), a lactose analogue which can activate promoters derived from the lac operon, was used to induce pGEX-4T-1-*CDKAL1* gene expression. Briefly, pGEX-T4-1-*CDKAL1* constructs were transformed into the protease deficient *E.coli* BL21 strain. The following day, a representative colony was inoculated in 4 mL LB supplemented with 50 μg/mL ampicillin. Two-hundred microlitres of overnight culture were grown for a further 15 hours in 4 mL of LB/ampicillin. A 200 μL aliquot of the culture was transferred to a 1.5 mL tube and grown in the absence of any stimulation (pre-induction sample). Eight microlitres 100 mM IPTG were added to the rest of the culture, which was grown for a further 3 hours at 37°C. Three further 200 μL aliquots were removed from the culture at 1 hour intervals (post-induction samples). Pre- and post induction cultures were centrifuged at 10,000 rpm for 5 min at room temperature and the pellet was re-suspended in 200 μL 1X Laemli buffer. Samples were denatured at 90°C and stored at -20°C or loaded on an 8% SDS-polyacrylamide minigel (Table 2.9.5.1). Samples were run at 120V for 1.5 hours, stained for 1 hour in comassium blue and then de-stained for an additional hour in 40% Methanol/10% Acetic acid. Gel images were acquired using a PowerShot A640 camera (Canon).

2.10 Statistical analyses

2.10.1 Power calculations

The statistical power of the UK case-control dataset was estimated based on published data on marker allele frequencies and odds ratios (Scott, Mohlke et al. 2007; Barrett, Hansoul et al. 2008). All calculations were carried out using the on-line Genetic Power Calculator (available at http://pngu.mgh.harvard.edu/~purcell/gpc/).

2.10.2 Genotype imputation

As *CDKAL1* SNP rs6908425 had not been analysed in the CASP study, genotypes were imputed using data from neighbouring SNPs. This analysis was carried out using PLINK 1.03 (Purcell, Neale et al. 2007) and results were confirmed with MACH (Li, Willer et al. 2010).

2.10.3 Analysis of case-control data

Chi-square tests were implemented on 2 x 3 contingency tables, in order to verify that genotype frequencies did not deviate from Hardy Weinberg Equilibrium (HWE). The 1958 Birth Cohort genotype database (http://www.b58cgene.sgul.ac.uk/) was queried in order to confirm that the allele frequencies of associated markers were homogenous across the UK.

Association analyses for the UK dataset were performed by combining the allele numbers of cases with those of the WTCCC controls (https://www.wtccc.org.uk/index.shtml) in 2 x 2 contingency tables. The genotypes generated by the CASP study (data available after registration at http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study), were analysed with PLINK 1.03 (Purcell, Neale et al. 2007). The meta-analysis of the UK and CASP data was carried out with Review manager 5 (http://ims.cochrane.org/revman/download).

2.10.4 Analysis of linkage disequilibrium conservation

The conservation of linkage disequilibrium (LD) within disease associated intervals was assessed using Haploview 3.32 (http://www.broad.mit.edu/mpg/haploview/). This software, which generates a graphic overview of the degree of LD between selected SNPs, was used to analyse the data generated by the HapMap consortium (Thorisson,

Smith et al. 2005) for the regions of interest.

2.10.5 Multiple testing corrections

The proportion of true association effects was assessed using two corrections for multiple testing. First, corrected p values (p_c) were calculated according to the Bonferroni formula ($p_c = p\ (n\text{-}1)$, where n is the number of analysed SNPs). Next, a False Discovery Rate (FDR) was calculated as follows: FDR= Nα/K, where N= total number of SNPs, α= threshold p value, K= number of SNPs where $p \leq \alpha$) (Benjamini 1995).

2.11 Bioinformatics

2.11.1 Evolutionary conservation analysis

The ClustalW2 software (www.ebi.ac.uk/Tools/clustalw2/index.html) was used to analyze the conservation of the putative CDKAL1 domains annotated in the Ensembl Genome browser (www.ensembl.org).

The UCSC (University of California Santa Cruz) genome Browser (http://genome.ucsc.edu/index.html) was queried, in order to assess the evolutionary conservation of non-coding disease-associated intervals. This analysis was based on the multiple alignment of 44 vertebrate genomes.

2.11.2 Analysis of microarray data

A gene expression profile was generated in quadruplicate for each sh-RNA construct. The readings generated by the iScan System (Illumina) were normalized to the 50thpercentile, using the GenomeStudio software. Normalized data were analyzed with the Qlucore Omics Explorer 2.1 software, using default options. The analysis was

performed in log2 scale and a t-test was used to generate P values. These were adjusted for multiple comparisons by calculating a FDR, using the formula described above (Benjamini 1995; Smyth 2004). Gene ontology (GO) analysis, was performed using GeneGo MetaCore pathway analysis version 6.4 (GeneGo Inc.), based on the GO terms annotated in the Ensembl genome browser. This analysis allows the reconstruction of putative cellular pathways, based on the mining of literature databases and the expression levels of the analyzed genes. Statistical significance is derived based on the overlap between the analyzed dataset and the genes participating to specific cellular pathways. Gene expression changes were expressed as log2 ratios.

Array data from the CASP were obtained from the GAIN database, through dbGaP accession number phs000019.v1.p1. Raw data were subjected to quantile normalization and expression estimates were computed using the Robust Multichip Average (RMA) method (Irizarry, Hobbs et al. 2003). z-scores were calculated for the differentially expressed genes, using the formula: $z = \frac{x_i - \bar{x}}{sd.}$, where x_i is the expression value of the individual being analyzed, $\bar{x}$ is the mean expression value of healthy individuals and sd the standard deviation. A set of dissimilarities was extracted for each gene and hierarchical clustering analysis was implemented using the *R* software (freely available at http://www.R-project.org).

2.11.3 Analysis of real-time PCR data

The expression of transcripts of interest was normalized to that of the *PPIA* (encoding cyclophilin A) and *RPLPO* (encoding large PO ribosomal protein) housekeeping genes (Table 2.11.3.1). Data analysis was performed using the ΔΔCycles threshold (Ct) method, which does not require the use of a standard curve (Livak and Schmittgen

2001; Lopez-Albaitero, Mailliard et al. 2009). The ΔΔCt value was calculated using the formula: $\Delta\Delta Ct = (Ct_{GOI} - Ct_{HK})$, where Ct_{GOI} and Ct_{HK} are the averaged Ct values for the gene of interest and the housekeeping gene. Data were expressed as mRNA expression fold changes, relative to a calibrator sample. Assuming a doubling of material during each PCR cycle, this relative quantification (RQ) was estimated according to the formula: $(RQ) = 2^{-\Delta\Delta Ct}$.

Table 2.11.3.1: TaqMan assays used in this study

Assay ID	Gene Symbol	Interrogated Sequence	Exon Boundary	Amplicon length (bp)
Hs99999904_m1	*PPIA*	NM_021130.3	4 - 4	98
Hs99999902_m1	*RPLP0*	NM_053275.3	3 - 3	105
Hs00214949_m1	*CDKAL1*	NM_017774.2	10 - 11	92
Hs00221754_m1	*TRIB3*	NM_021158.3	2-3	115
Hs01090850_m1*	*DDIT3*	NM_004083.4	1-2	78
Hs00326824_m1*	*RAB3GAP1*	D31886.1	23-24	85
Hs00207275_m1*	*RIMS3*	NM_014747.2	5-6	57
Hs00368884_g1	*INHBE*	NM_031479.3	1-2	68
Hs00262772_m1	*LSM10*	NM_032881.1	1-2	113
Hs01595873_m1*	*GTF2H5*	NM_207118.2	2-3	138
Hs00897129_m1	*NME1*	NM_198175.1	2-3	72
Hs00157694_m1*	*GCLM*	NM_002061.2	1-2	82
Hs00607181_g1*	*TUBB2C*	NM_006088.5	3-4	138
Hs00988959_gH	*RPL29*	NM_000992.2	4-4	61

Hs00249482_m1*	*SEPX1*	NM_016332.2	2-3	67
Hs00360764_m1	*POLR2L*	U37690.1	1-2	74
Hs00733231_m1*	*RPL36A*	NM_001001.3	1-2	111

Chapter 3: Genetic analysis of 26 Crohn's disease (CD) susceptibility loci in psoriasis

3.1 Association analysis

An association study was undertaken, in order to investigate the possibility of a genetic overlap between psoriasis and CD. At the onset of this project, GWASs had identified 32 loci (including *IL12B* and *IL23R*) significantly associated with CD ($p < 10^{-5}$) (Parkes, Barrett et al. 2007; The Wellcome Trust Case-Control Consortium 2007; Barrett, Hansoul et al. 2008). Power calculations demonstrated that the dataset available for this study had adequate power to detect associations at 28 CD loci (Table 3.1.1). On this basis, a genotyping panel (Table 3.1.2) was designed, which included one marker from each of the relevant CD susceptibility regions. Two markers were typed from the 1q24 (rs12035082 and rs9286879) and 5q33 (rs11747270 and rs1000113) loci, based on reports of associations at two unrelated SNPs ($r^2 < 0.4$) (Parkes, Barrett et al. 2007; The Wellcome Trust Case-Control Consortium 2007; Barrett, Hansoul et al. 2008). No susceptibility alleles from the *CARD15*/*NOD2* gene were analyzed, as previous studies had failed to identify any evidence for an involvement of these variants in psoriasis (Nair, Stuart et al. 2001; Borgiani, Vallo et al. 2002; Young, Allen et al. 2003). Moreover, CD markers lying on chromosome 6p21 were not analysed, as the linkage disequilibrium with *PSORS1* alleles would have confounded the interpretation of association signals.

The SNPs listed in Table 3.1.1 were typed in 1,256 psoriatic patients and the resulting allele frequencies were compared with those generated by the Wellcome Trust Case Control Consortium (WTCCC) for 2,938 controls from the 1958 British Birth Cohort and the National Blood Service (The Wellcome Trust Case-Control Consortium 2007). The use of control genotypes generated in another centre was validated by typing one

randomly selected SNP (rs12035082) in 1,304 individuals from the 1958 British Birth Cohort. No discrepancies between these genotypes and the ones from the WTCCC study were observed. At the same time, the quality of patient genotype data was assessed by investigating deviations from Hardy-Weinberg Equilibrium (HWE). This identified a significant departure from HWE for SNP rs11175593 ($p = 0.0002$). A re-analysis of the SNP data, based on the direct sequencing of 192 patient samples, revealed a number of inaccuracies in the results obtained by TaqMan genotyping. On this basis, marker rs11175593 was excluded from all subsequent analyses.

The comparison of case and control allele frequencies identified significant disease associations for four independent markers: rs12035082 ($p = 0.009$; OR = 1.14; 95% CI: 1.03-1.25), rs6908425 ($p = 0.00015$; OR = 1.26; 95% CI: 1.12-1.42), rs2836754 ($p = 0.0003$; OR = 1.15; 95% CI: 1.06-1.30) and rs762421 ($p = 0.0002$; OR = 1.22; 95% CI: 1.09-1.36). Moreover, associations of marginal significance were identified for SNPs rs1000113 ($p = 0.045$; OR = 1.20; 95% CI: 1.00-1.45) and rs4263839 ($p = 0.04$; OR = 0.9; 95% CI: 0.81-0.99).

In order to address the issue of multiple testing, a Bonferroni correction was applied to p values and a False Discovery Rate (FDR) was calculated. Following these analyses, four markers (listed in Table 3.1.3) remained associated at a FDR < 0.01 and 3 yielded corrected p values < 0.05 (Table 3.1.3). To exclude the possibility that these results may reflect the presence of population stratification, the control genotypes of the 1958 British Birth cohort were examined in more detail. Mining of a public database (available at http://www.b58cgene.sgul.ac.uk/) showed that all 3 critical SNPs had homogeneous allele frequencies across UK regions, thereby excluding the possibility that the associations may be due to the occurrence of hidden population structure. Finally, a logistic regression analysis was carried out, using age and sex as covariates.

No increase in the significance of p values was observed, indicating that the associations described here are not influenced by either of these variables.

Table 3.1.1: Power calculations

				N. cases for 80% power	
Marker	**OR**	**Chr.**	**Gene***	**p = 0.05**	**p = 0.01**
rs12035082	1.14	1q24	-	245	364
rs9286879	1.19	1q24	-	931	1,385
rs10801047	1.47	1q31	-	433	644
rs10210302	1.19	2q37	*ATG16L1*	731	1,088
rs9858542	1.17	3p21	*MST1*	1,043	1,553
rs17234657	1.16	5p13	-	219	327
rs6596075	1.55	5q31	*SLC22A4*	164	245
rs11747270	1.33	5q33	*IRGM*	780	1,161
rs1000113	1.54	5q33	*IRGM*	344	512
rs3763313	1.19	6p21	*Many*	904	1,346
rs6908425	1.21	6p22	*CDKAL1*	957	1,424
rs7746082	1.17	6q21	-	1,021	1,506
rs6927210	1.13	6q23	-	648	965
rs2301436	1.21	6q27	*CCR6*	658	980
rs1456893	1.18	7p12	-	713	1,062
rs887822	1.20	7q36	-	233	347
rs7849191	**1.12**	**9p24**	***JAK2***	**1,849**	**2,752**
rs4263839	1.22	9q32	*TNFSF15*	528	786
rs17582416	1.16	10p11	-	997	1,483
rs6601764	1.16	10p15	-	1,020	1,518
rs10761659	1.23	10q21	*ZNF365*	266	396
rs10883365	1.18	10q24	*NKX2-3*	770	1,147
rs7927894	1.16	11q13	*C11orf30*	1,025	1,526
rs11175593	1.54	12q12	*MUC19*	1,291	1,922
rs2066847	3.99	16q12	*NOD2/CARD15*	66	98
rs744166	1.18	17q21	*STAT3*	797	1,187
rs2872507	**1.12**	**17q21**	***ORMDL3***	**1,610**	**2,396**
rs2542151	1.35	18p11	*PTPN2*	368	548

rs4807569	1.02	19p13	*SBNO2*	1,294	1,926
rs1736135	1.18	21q21	-	822	1,224
rs2836754	1.15	21q22	*FLJ45139*	1,141	1,698
rs762421	1.13	21q22	*ICOSLG*	618	920

Calculations were carried out, based on the risk allele frequencies and odds ratios (OR) observed in published CD GWASs, assuming the analysis of 2,938 controls. Markers that were excluded from further analyses, due to lack of power are highlighted in bold. Markers from the *IL23R* and *IL12B* loci were not included in the power calculations or in further analyses, as associations with psoriasis had already been established. * Gene of interest found in the region tagged by the examined SNP.

Table 3.1.2: Association analysis results

			Allele Counts (Minor Allele Frequency)			
Marker	**Chr.**	**Gene***	**Cases**	**Controls**	**OR**	**p value**
rs12035082	**1q24**	**-**	**0.42 (993/2366)**	**0.39 (2281/5868)**	**1.14**	**0.009**
rs9286879	1q24	-	0.26 (502/1946)	0.24 (1374/5776)	1.07	0.07
rs10801047	1q31	-	0.07 (164/2338)	0.07 (395/5860)	1.04	0.66
rs10210302	2q37	*ATG16L1*	0.48 (1157/2386)	0.48 (2822/5872)	1.02	0.71
rs9858542	3p21	*MST1*	0.29 (692/2360)	0.28 (1652/5862)	1.06	0.30
rs17234657	5p13	-	0.13 (319/2448)	0.12 (731/5876)	1.05	0.47
rs6596075	5q31	*SLC22A4*	0.17 (396/2362)	0.17 (972/5870)	0.98	0.82
rs11747270	5q33	*IRGM*	0.07 (138/1850)	0.07 (414/5876	1.06	0.54
rs1000113	5q33	*IRGM*	0.08 (183/2288)	0.07 (394/5854)	1.20	0.045
rs6908425	**6p22**	***CDKAL1***	**0.19 (452/2358)**	**0.23 (1348/5864)**	**0.79**	**0.00015**
rs7746082	6q21	-	0.30 (645/2176)	0.29 (2475/8656)	1.05	0.33
rs6927210	6q23	-	0.50 (1200/2388)	0.48 (2825/5866)	1.09	0.08
rs2301436	6q27	*CCR6*	0.49 (892/1838)	0.48 (2749/5776)	1.03	0.48
rs1456893	7p12	-	0.33 (609/1842)	0.32 (2817/8754)	0.96	0.46
rs887822	7q36	-	0.27 (655/2384)	0.29 (1695/5868)	0.93	0.20
rs4263839	9q32	*TNFSF15*	0.31 (710/2278)	0.33 (2936/8784)	0.90	0.04
rs17582416	10p11	-	0.33 (591/1814)	0.33 (1943/5878)	1.02	0.58
rs6601764	10p15	-	0.42 (973/2316)	0.41 (2368/5810)	1.07	0.20
rs10761659	10q21	*ZNF365*	0.47 (1103/2326)	0.46 (2703/5866)	1.05	0.27
rs10883365	10q24	*NKX2-3*	0.49 (1081/2218)	0.47 (2730/5800)	1.07	0.18
rs7927894	11q13	*C11orf30*	0.39 (669/1702)	0.40 (2332/5872)	0.99	0.7
rs744166	17q21	*STAT3*	0.44 (990/2266)	0.44 (1251/2854)	0.98	0.7
rs2542151	18p11	*PTPN2*	0.17 (372/2140)	0.16 (957/5870)	1.08	0.25

rs4807569	19p13	*SBNO2*	0.21 (402/1880)	0.21 (1205/5874)	0.94	0.4
rs1736135	21q21	-	0.42 (964/2298)	0.42 (1245/2960)	0.96	0.57
rs2836754	**21q22**	***FLJ45139***	**0.39 (847/2168)**	**0.35 (2072/5862)**	**1.17**	**0.0003**
rs762421	**21q22**	***ICOSLG***	**0.34 (655/1900)**	**0.39 (2303/5876)**	**0.82**	**0.0002**

*Gene found in the susceptibility interval tagged by the examined SNP. OR: odds ratio for minor allele. The bold font denotes markers generating a FDR < 0.01.

Table 3.1.3: Association results for SNPs generating a FDR < 0.01

			Minor allele frequency				
Marker	**Chr.**	**Gene***	**Cases**	**Controls**	**p value**	**p_c value**	**OR**
rs12035082	1q24	-	0.42	0.39	0.009	0.24	1.14
rs6908425	**6p22**	***CDKAL1***	**0.19**	**0.23**	**0.00015**	**0.004**	**0.79**
rs2836754	**21q22**	***FLJ45139***	**0.39**	**0.35**	**0.0003**	**0.008**	**1.17**
rs762421	**21q22**	***ICOSLG***	**0.34**	**0.39**	**0.0002**	**0.005**	**0.84**

* Gene found in the susceptibility interval tagged by the examined SNP. OR: odds ratio for the minor allele; p_c value: p value following Bonferroni correction. The bold font denotes markers generating a $p_c < 0.05$.

3.2 Characterization of three psoriasis associated loci

The genotype data generated by the HapMap Consortium (Frazer, Ballinger et al. 2007) were analyzed, with a view to defining the genomic regions tagged by the three disease associated SNPs. The results of this analysis are shown in Figures 3.2.1, 3.2.2 and 3.2.3.

rs6908425 is a non-coding C/T SNP, which lies on chromosome 6p22.3, within intron 5 of the *CDKAL1* gene. The SNP is located in close proximity of a 48 kb LD block, which includes one SNP (rs10946398) previously associated with type 2 diabetes (T2D) (Saxena, Voight et al. 2007; Scott, Mohlke et al. 2007; Zeggini, Weedon et al. 2007). Although SNPs rs6908425 and rs10946398 lie in the same *CDKAL1* intron, they are not in LD with each other ($r^2 = 0.04$) (Fig. 3.2.1).

rs2836754 is a non-coding C/T SNP, which lies in intron 2 of the *FLJ45139* anonymous transcript, on chromosome 21q22.2. The SNP maps to a 3 kb LD block, which is delimited by markers rs8129427 (r^2 with rs2836754: 0.02) and rs2410067 (r^2 with rs2836754: 0.03) (Fig. 3.2.2). This block also contains several genomic segments that are highly conserved among mammals.

rs762421 is a non-coding G/A SNP, which lies on chromosome 21q22.3, in a gene desert, upstream of the *ICOSLG* transcript. The SNP maps to a 19 kb LD block, which is delimited by markers rs9976849 (r^2 with rs762421: 0.21) and rs2838528 (r^2 with rs762421: 0.09) (Fig. 3.2.3).

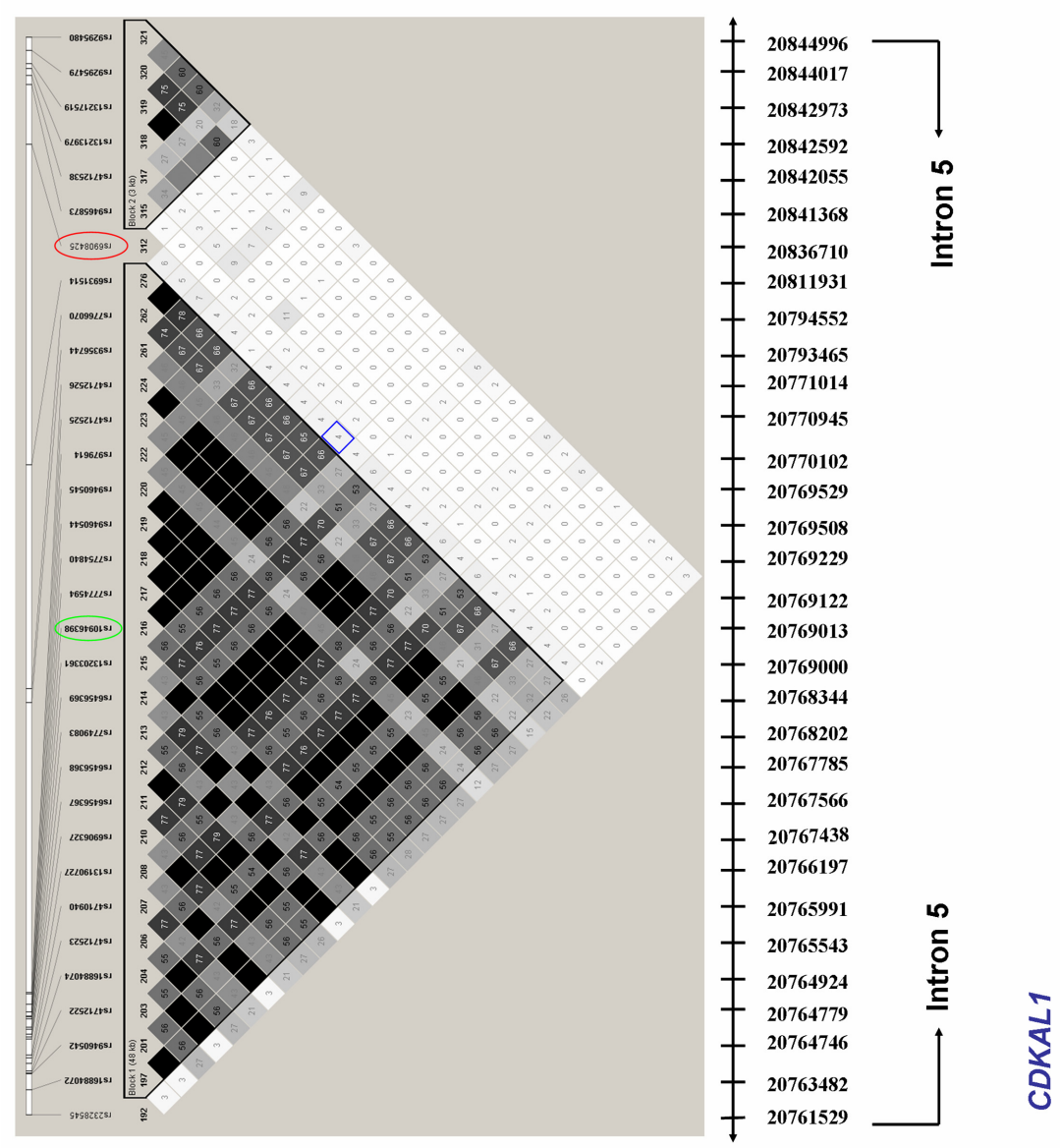


Figure 3.2.1: Overview of linkage disequilibrium conservation around SNP rs6908425

The position of marker rs6908425 is indicated by a red circle. The position of the T2D susceptibility marker (rs10946398) is indicated by a green circle. The blue square highlights the lack of correlation between rs6908425 and rs10946398 (r^2: 0.04). The diagram below the LD plot indicates the position of the investigated region with respect to chromosome 6 coordinates and *CDKAL1* intron 5.

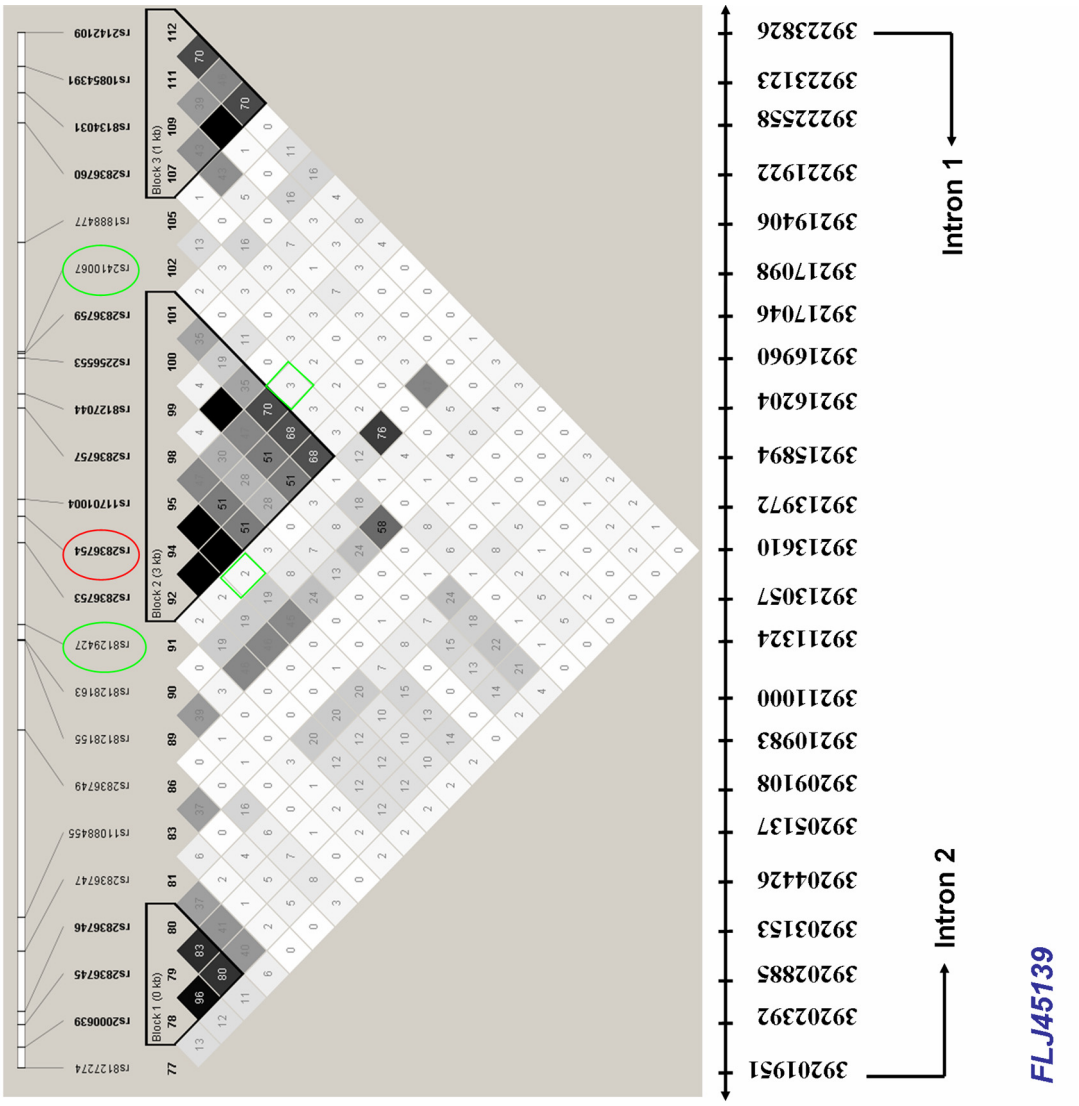


Figure 3.2.2: Overview of linkage disequilibrium conservation around SNP rs2836754

SNP rs2836754, which is indicated by a red circle, lies in a 3 kb LD block, delimited by markers rs8129427 and rs2410067 (green circles). Green squares highlight the r^2 between rs2836754 and rs8129427 (r^2: 0.02) and that between rs2836754 and rs2410067 (r^2: 0.03). The arrow below the LD plot indicates the position of the investigated region with respect to chromosome 21 coordinates and the genomic organization of *FLJ45139*.

rs28385831
rs9983276
rs9977351
rs7280525
rs2329711
rs2779626
rs28385281
rs12626606
rs3804031
rs28385826
rs28385824
rs28385822
rs2782137
rs6518350
rs8129605
rs2032327
rs2070551
rs4380337
rs28385821
rs28385820
rs2782490
rs762422
rs762421
rs28385819
rs28385818
rs2776302
rs28385817
rs9976849
rs743479
rs8740171
rs762418
rs762417
rs743477
rs743476
Block 2 (19 kb)
Block 1 (3 kb)
152
151
149
148
143
142
139
138
131
128
125
120
115
112
111
109
107
106
105
104
103
102
101
99
98
96
95
94
93
91
89
88
87
86
444634797
444349446
444352449
444353521
444356518
444363781
444378331
444382533
444385871
444392301
444394511
444399891
444400661
444401691
444403241
444403451
444404771
444411001
444417561
444423061
444462451
444490951
444523041
444528291
444537961
444540321
444576781
444579051
444600441
444608151
444618341
444618201
444627161
444630141

Figure 3.2.3: Overview of linkage disequilibrium conservation around SNP rs762421

SNP rs762421, which is indicated by a red circle, lies in a 19 kb LD block, delimited by markers rs9976849 and rs2838528 (green circles). Dark green squares highlight the r^2 between rs762421 and rs9976849 (r^2: 0.21) and that between rs762421 and rs2838528. The arrow below the LD plot indicates the position of the investigated region with respect to chromosome 21 coordinates.

3.3 Replication analysis of disease associated SNPs

To validate the observed disease associations, the genome-wide data generated by the Collaborative Association Study of Psoriasis (CASP) (Nair, Duffin et al. 2009) was queried. The study was based on the analysis of 1,348 cases and 1,368 controls and good quality SNP genotypes (call rate > 90%, p value for HWE > 0.05) were available for psoriasis associated markers rs2836754 (chr. 21q22) and rs762421 (chr. 21q22). As SNP rs6908425 (chr. 6p22) was not present in the Perlegen array used to generate the CASP data, the genotypes for this marker were imputed, based on those of neighbouring SNPs.

The results of the CASP dataset association analysis are summarized in Table 3.3.1.

A p value of 0.06 was observed for SNP rs2836754, at the *FLJ45139* locus. The analysis of neighbouring SNPs identified significant associations ($p < 0.01$) for at least three markers (rs2142107, rs13047416 and rs2836773), but an examination of LD conservation patterns failed to reveal any correlation between rs2836754 and the relevant SNPs. In fact the r^2 values between rs2836754 and rs2142107 (yielding a p of 0.0086), rs13047416 ($p = 0.0086$) and rs2836773 ($p = 0.002$) only range between 0.24 and 0.46 (Fig. 3.3.1).

Neither rs762421 nor any of its neighbouring SNPs in the *ICOSLG* gene region showed any significant association with psoriasis.

Finally, the comparison of case and control imputed allele frequencies for *CDKAL1* SNP rs6908425 generated a p value equal to 0.00012.

In the subsequent phase of the study, combined p values were calculated for the three critical markers, through a meta-analysis of the UK and CASP datasets. The results are reported in Table 3.3.1.

The significance of the association for SNP rs2836754 (*FLJ45139* locus) decreased in

the overall dataset, compared to the UK sample.

The marker lying in the *ICOSLG* gene region (rs762421) could not be analysed across the UK and CASP datasets, because of evidence of heterogeneity between the two samples ($p = 0.01$; $I^2 = 83.9\%$). Thus, only *CDKAL1* SNP rs6908425 generated a convincing association in both the discovery and the replication datasets. No evidence of heterogeneity was detected between the two samples ($p= 0.85$; $I^2 = 0\%$) and a meta-analysis of the two studies yielded a combined p value of 4×10^{-6} (OR: 1.26; 95% CI: 1.15–1.41) (Table 3.3.1).

Table 3.3.1: Analysis of psoriasis associated SNPs in the CASP dataset

Marker	Gene	p value (UK)	p value (CASP)	p value (combined)	OR (combined)
rs6908425	*CDKAL1*	0.00015	0.00012*	$4x10^{-6}$	0.79 (CI: 1.15-1.41)
rs2836754	*FLJ45139*	0.0003	0.06	0.001	1.15 (CI: 1.06-1.25)
rs762421	*ICOSLG*	0.0002	0.50	N/A**	

*based on imputed genotypes; ** N/A: Non applicable, as the meta-analysis uncovered evidence for heterogeneity between the two samples. OR: odds ratio for the minor allele; CI: 95% confidence intervals.

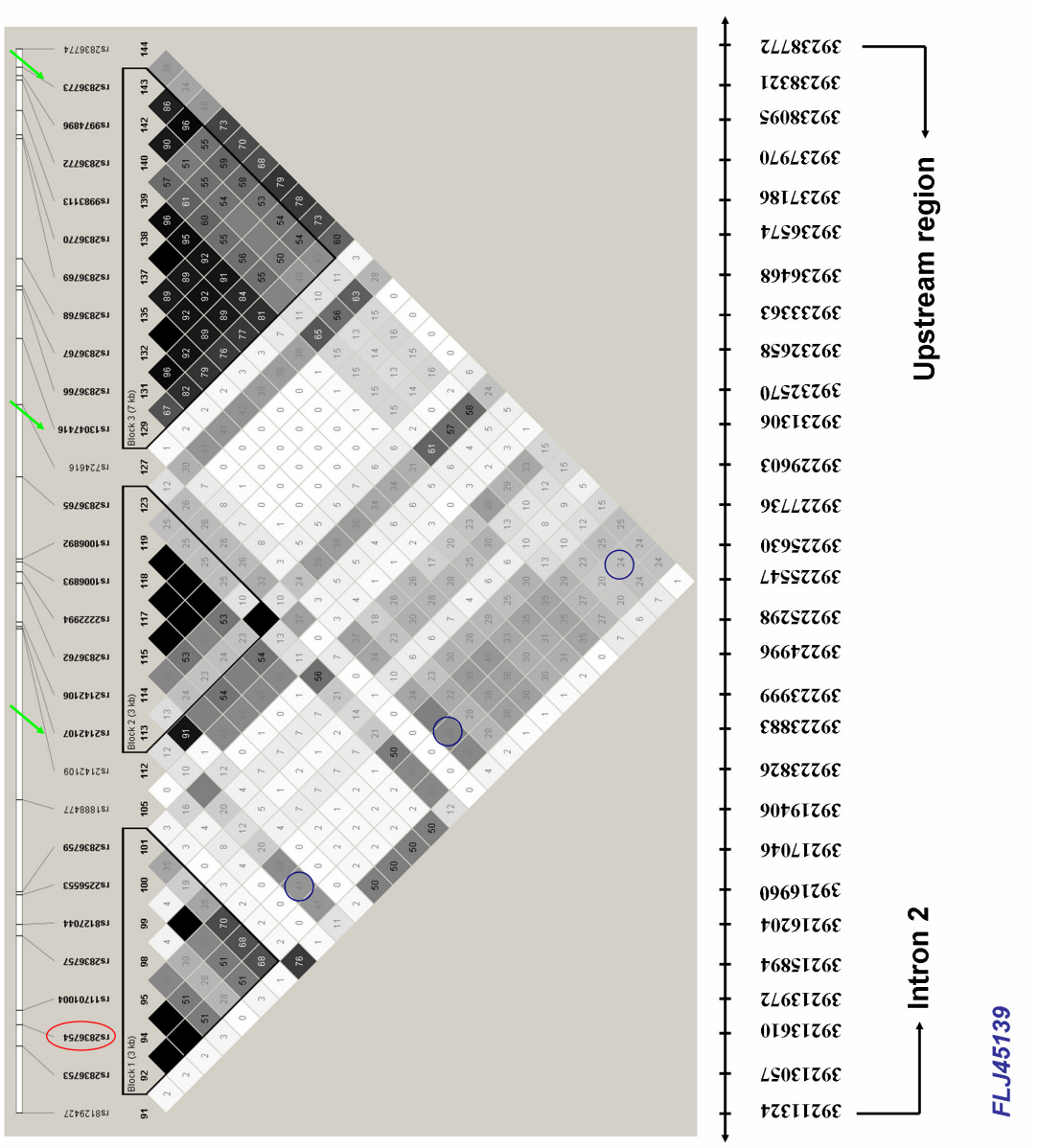


Figure 3.3.1: Position of linkage disequilibrium blocks around SNP rs2836754

SNP rs2836754, indicated by a red circle, lies in a 3 kb LD block. Markers rs2142107, rs13047416 and rs2836773 (indicated by green arrows), which have shown association in the CASP study, map to different LD blocks. Blue circles indicate the correlation between rs2836754 and rs2142107 (r^2: 0.41); rs2836754 and rs13047416 (r^2: 0.46); rs2836754 and rs2836773 (r^2: 0.24). The arrow below the LD plot indicates the position of the investigated region with respect to chromosome 21 coordinates and the genomic organization of *FLJ45139*.

3.4 Analysis of allelic heterogeneity at the *CDKAL1* locus

The analysis of the data generated by the CASP project for the *CDKAL1* locus identified significant disease associations ($p < 0.01$) for several SNPs. Of note, many of these markers were not correlated with rs6908425 ($r^2 \leq 0.05$), suggesting the presence of allelic heterogeneity. To further investigate this possibility, four representative SNPs, unrelated to each other and to rs6908425, were selected for follow-up in the UK dataset (Fig. 3.4.1 and Table 3.4.1). These SNPs were genotyped in 962 UK psoriatic cases and the resulting allele frequencies were compared to those of the 2,938 unrelated controls analyzed by the WTCCC. Following the quality control of genotype data, a significant departure from HWE ($p = 0.0007$) was identified for one of the examined SNPs (rs4454125), which was not analyzed any further. Despite evidence for power (Table 3.4.1), no significant association with psoriasis was identified for any of the remaining markers (Table 3.4.2).

Intron 13
Intron 3
CDKAL1

Figure 3.4.1: Overview of linkage disequilibrium conservation across the *CDKAL1* gene

The position of marker rs6908425 is indicated by a red circle. The positions of SNPs showing association in the CASP study are indicated by blue arrows. The LD plot shows that psoriasis associated SNPs rs714831, rs7752602, rs4454125 and rs714830 are independent from rs6908425, as shown by the r^2 values highlighted by blue circles. The arrow below the LD plot indicates the position of the investigated region with respect to chromosome 6 coordinates and the genomic organization of *CDKAL1*.

Table 3.4.1: Power calculations for SNPs rs714831, rs714830, rs4454125 and rs775202

	N. cases for 80% power	
Marker	**p = 0.05**	**p = 0.01**
rs714831	660	982
rs714830	663	987
rs4454125	920	1,398
rs7752602	575	855

Calculations were carried out, based on the minor allele frequencies and odds ratios observed in the CASP dataset and assuming the analysis of 2,938 controls.

Table 3.4.2: Analysis of allelic heterogeneity at the *CDKAL1* locus

		Allele Counts (Minor Allele Frequency)			
Marker	**Position**	**Cases**	**Controls**	**p value (UK)**	**p value (CASP)**
rs714831	Intron 4	0.50 (943/1900)	0.48 (2846/5876)	0.36	0.001
rs714830	Intron 4	0.13 (248/1844)	0.13 (775/5878)	0.78	0.0006
rs4454125	Intron 12			Out of HWE	0.005
rs7752602	Intron 13	0.32 (603/1868)	0.33 (1951/5876)	0.55	0.0004

3.5 Analysis of type 2 diabetes susceptibility SNPs in the psoriasis dataset

GWASs have identified associations with T2D for several correlated markers lying within *CDKAL1* intron 5 (rs10946398, rs9465871, rs7767391) (Scott, Mohlke et al. 2007; Zeggini, Weedon et al. 2007). To explore the relationships between the T2D and psoriasis association signals, the data generated by the HapMap consortium for the *CDKAL1* region was queried. This analysis revealed a complete lack of correlation between psoriasis and T2D associated markers, as the r^2 values between rs6908425 and rs10946398, rs9465871, rs9460546 were all ≤ 0.04 (Fig. 3.5.1). This observation indicates that the two association signals are unrelated, but does not exclude the possibility that T2D SNPs may confer susceptibility to psoriasis, independently from rs6908425. To better investigate this hypothesis, a case-control study was carried out for the marker showing the strongest association with T2D (rs10946398). This SNP was genotyped in 962 UK psoriatic cases and the resulting allele frequencies were compared to those of the 2,938 unrelated controls analyzed by the WTCCC. Although the dataset had adequate statistical power to detect an association with rs10946398 (Table 3.5.1), none was observed ($p = 0.60$).

Finally, the relationship between the CD and the T2D associations was examined. For this purpose, the data generated by the WTCCC for CD (The Wellcome Trust Case-Control Consortium 2007) was queried, to assess the presence of a disease association with the T2D marker. Despite the adequate power of the dataset, which included 1,747 patients and 2,938 controls (Table 3.5.1), no association was detected between rs10946398 and CD ($p = 0.84$).

Taken together, these data support the presence of an allelic series at the *CDKAL1* locus, where unrelated SNPs confer susceptibility to psoriasis, CD (rs6908425) and, T2D (rs10946398).

Intron 13
Intron 3
CDKAL1
Block1 (48 kb)

Figure 3.5.1: Overview of linkage disequilibrium conservation across the *CDKAL1* locus

The position of marker rs6908425 is indicated by a red circle. The positions of SNPs showing association with T2D are indicated by green arrows. The LD plot shows that the T2D susceptibility SNPs are independent from rs6908425, as shown by the r^2 values highlighted by dark green circles. The arrow below the LD plot indicates the position of the investigated region with respect to chromosome 6 coordinates and the genomic organization of *CDKAL1*.

Table 3.5.1: Power calculations for SNP rs10946398

	N. cases for 80% power		
Marker	**p = 0.05**	**p = 0.01**	**OR**
rs10946398	633	943	1.2

Calculations were carried out, based on the minor allele frequencies and odds ratios observed in the WTCCC dataset, and assuming the analysis of 2,938 controls.

3.6 Discussion

The existence of shared genetic determinants for clinically distinct inflammatory disorders is clearly documented by the results emerging from several GWASs (Seldin and Amos 2009). In this context, the aim of this study was to specifically investigate the genetic overlap between psoriasis and CD, based on the observation that at least two genes (*IL23R* and *IL12B)* harbor alleles conferring susceptibility to both conditions (Duerr, Taylor et al. 2006; Capon, Di Meglio et al. 2007; Cargill, Schrodi et al. 2007; Barrett, Hansoul et al. 2008). To achieve this purpose, an association analysis was undertaken, by genotyping 28 markers that had been associated with CD in GWASs carried out prior to the onset of this study (Parkes, Barrett et al. 2007; Barrett, Hansoul et al. 2008). This identified three CD markers (rs6908425, rs2836754 and rs762421) that were associated with psoriasis at an FDR < 0.01. Given the evidence for adequate statistical power in the study dataset, the lack of association at the other loci is likely to reflect genuine differences in the genetic architecture of psoriasis and CD.

The analysis of an independently ascertained case-control dataset (the CASP study) validated the association for *CDKAL1* marker rs6908425. The lack of association at the two remaining loci may reflect the lower statistical power of the CASP dataset (including 1,348 cases and 1,368 controls) compared with the UK resource (1,256 cases and 2,938 WTCCC controls).

Of interest, a number of *CDKAL1* markers unrelated to rs6908425 were found to be significantly associated with psoriasis in the CASP study. However, the analysis of these SNPs in the UK dataset failed to reveal any evidence of association, with one marker also showing a significant departure from HWE. These observations suggest that these particular CASP findings are likely to represent false positive results, possibly reflecting the difficulty of genotyping some *CDKAL1* markers.

Marker rs6908425 maps in intron 5 of *CDKAL1,* which also harbours a cluster of SNPs that have been repeatedly associated with T2D (Saxena, Voight et al. 2007; Scott, Mohlke et al. 2007; Steinthorsdottir, Thorleifsson et al. 2007; Zeggini, Weedon et al. 2007). T2D is a common condition and its frequency in the UK is approximately 5-10% (Penn, White et al. 2009). There is also clinical evidence that psoriatic patients are at higher risk to develop T2D than the general population (Gottlieb and Dann 2009). Thus, a significant minority of the patients from the UK and CASP datasets are likely to be affected by T2D, as well as psoriasis. Although this is undoubtedly a confounding factor, the possibility that the association seen in the UK and CASP cohorts could be solely determined by the subset of patients with T2D is unlikely, given the poor statistical power afforded by this subgroup.

To further explore the relationships between the T2D and psoriasis association signals, a *CDKAL1* SNP associated to T2D (rs10946398) was analysed in the UK psoriasis dataset. Despite the adequate statistical power of this sample, no evidence for association was observed. Likewise, an analysis of the data generated by the WTCCC failed to reveal any evidence of association between rs10946398 and CD.

Taken together, the results presented in this chapter identify *CDKAL1* as a novel psoriasis susceptibility gene and indicate the presence of an allelic series, whereby distinct intron 5 SNPs confer susceptibility to inflammatory (CD, psoriasis) and metabolic disorders (T2D).

Chapter 4: Analysis of *CDKAL1* gene expression

4.1 The *CDKAL1* locus

CDKAL1 lies on chromosome 6p22.3, where it spans a 697,948 base pair (bp) region, extending from position 20,534,688 to 21,232,635, in the GRCh37 human genome assembly. Five *CDKAL1* isoforms are annotated in the Ensembl genome browser (release 56): *CDKAL1*-201, *CDKAL1*-202, *CDKAL1*-203, *CDKAL1*-001 and *CDKAL1*-002. These transcripts differ in the length of their untranslated regions (UTRs) and in the number of coding exons (Fig 4.1.1). *CDKAL1*-201 measures 2,408 bp and represents the full length transcript. It consists of 16 exons and encodes a 579 amino acid (Aa) protein that has a predicted molecular weight of 65 kDa. *CDKAL1*-202 spans 1,784 bp and lacks exons 4 and 13, as well as the final portion of the gene 3' UTR. It encodes a 488 Aa protein with a predicted molecular weight of 55 kDa. *CDKAL1*-203 includes all 16 exons, but only measures 1,960 bp, as its 3' UTR is shorter than that of the full length mRNA. *CDKAL1*-001 spans 3,115 bp; it lacks non-coding exons 1 and 2, but includes a longer 3' UTR region, compared to the full length transcript. *CDKAL1*-002 is 840 bp in length and its sequence includes two non-coding exons lacking an ATG codon (Fig. 4.1.1).

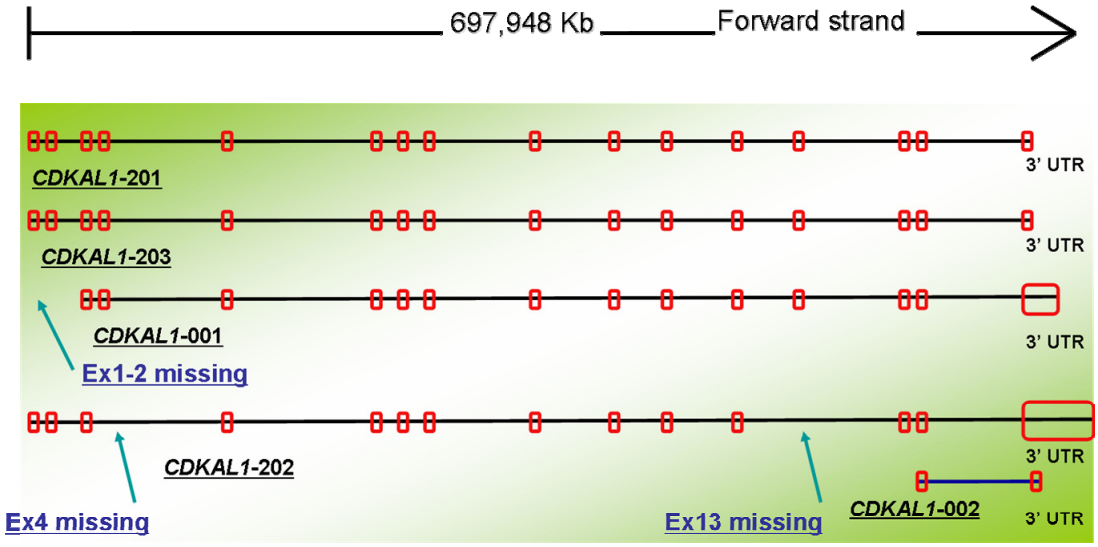


Figure 4.1.1: Graphic representation of *CDKAL1* isoforms

Black and blue bars represent coding and non-coding transcripts, respectively. Red squares indicate the position of the various exons. Ex: exon.

Adapted from the transcript diagram available in the Ensembl genome browser.

4.2 Experimental analysis of *CDKAL1* transcript isoforms

CDKAL1-202 is the only isoform that is predicted to encode a smaller protein product than the one generated by the full length transcript. For this reason, an RT-PCR analysis was carried out, to specifically investigate *CDKAL1*-202 tissue distribution. RT-PCR primers annealing to exon one and sixteen were designed, in order to amplify the full length cDNA as well as transcripts lacking exons four and thirteen. The primers were used to amplify a panel of 11 cDNAs derived from a variety of tissues, including several that are affected by T2D (skeletal muscle, pancreas), CD ($CD4^+$ and $CD8^+$ T lymphocytes, colon) and psoriasis ($CD4^+$ and $CD8^+$ T lymphocytes, skin keratinocytes). Electrophoresis analysis of the RT-PCR products showed the amplification of a single band, in all examined tissues (Fig.4.2.1). The molecular weight of the band was consistent with that of the full-length cDNA and sequence analysis confirmed that the RT-PCR product was indeed *CDKAL1*-201. Thus the *CDKAL1*-202 isoform was not detected in any of the examined tissues.

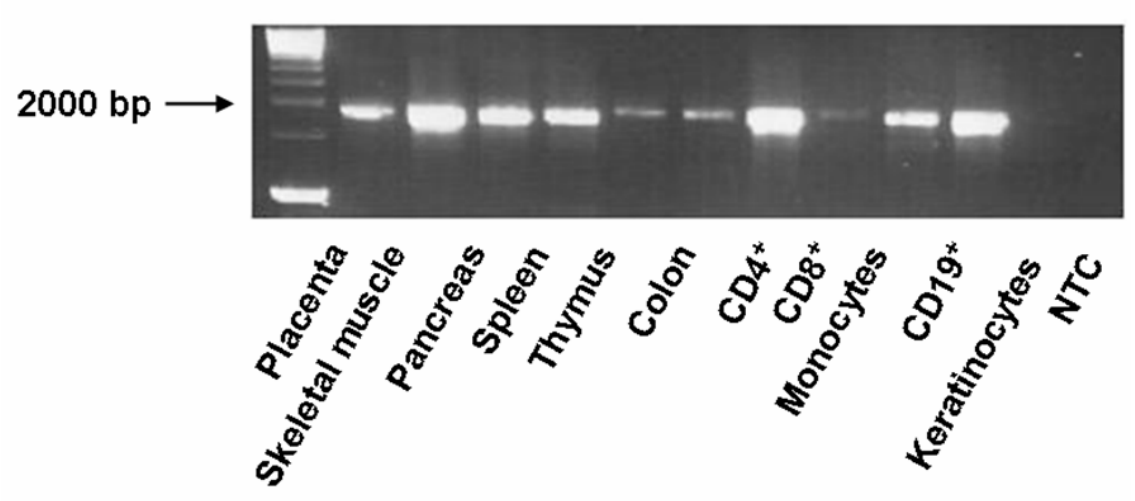


Figure 4.2.1: Electrophoresis analysis of *CDKAL1* RT-PCR products

A single amplification product consistent with the size of the full-length cDNA (1,960 bp, based on the position of RT-PCR primers) was visualized on the gel. $CD4^+$: $CD4^+$ T lymphocytes; $CD8^+$: $CD8^+$ T lymphocytes; $CD19^+$: $CD19^+$ B lymphocytes; NTC: no template control.

4.3 Analysis of *CDKAL1* tissue distribution

The function of the CDKAL1 protein is currently unknown and only limited information is available on the gene expression pattern, as previous studies mainly focused on tissues that are relevant to the pathogenesis of T2D (Scott, Mohlke et al. 2007; Zeggini, Weedon et al. 2007). To address this issue, a comprehensive characterization of *CDKAL1* transcript levels was performed by real-time PCR, using a panel of cDNAs derived from 20 tissues and cell types. This analysis revealed a widespread gene expression pattern, with *CDKAL1* transcripts detectable in all examined tissues. In agreement with published data (Scott, Mohlke et al. 2007), very strong *CDKAL1* expression was observed in skeletal muscle, more than two-fold higher than in the brain. Weaker expression levels were also detected in the pancreas, liver and kidney (Fig.4.3.1). Surprisingly, in CD target tissues such as the small intestine and colon, the expression of *CDKAL1* was low. A similar expression pattern was observed in skin keratinocytes, a cell type proven to play a role in psoriasis pathogenesis. Conversely, *CDKAL1* was strongly expressed in immune cells, such as $CD4^+$ T lymphocytes, $CD8^+$ T lymphocytes, $CD19^+$ B lymphocytes and monocytes (Fig.4.3.1).

CDKAL1 expression was further investigated in lymphocytes and keratinocytes, as these are the main cell types contributing to the pathogenesis of psoriasis.

Cultured primary keratinocytes, obtained from skin biopsies of one healthy and one psoriatic donor, were treated with a panel of pro-inflammatory molecules (IFN-α, IFN-γ, TNF-α, IL-17, IL-22 and IL-23) and the relevant cDNAs were analyzed by real-time PCR. As shown in Fig.4.3.2A, no changes in *CDKAL1* expression were observed in normal primary keratinocytes, following treatment with any of the examined cytokines. However, a marked *CDKAL1* down-regulation was detected in psoriatic keratinocytes treated with IL-17, IL-23 and IFN-α (Fig. 4.3.2B).

To explore the role played by *CDKAL1* in lymphocytes, the gene expression was investigated in mitogen-activated T and B cells. Real-time PCR experiments were carried out on commercially available cDNAs, obtained from pools of resting and activated $CD4^+$, $CD8^+$ and $CD19^+$ lymphocytes. These cells had been purified from blood samples of healthy donors and had been activated with concavalin A ($CD4^+$ T lymphocytes), phytohaemoagglutinin ($CD8^+$ T lymphocytes) and pokeweed mitogen ($CD19^+$ B lymphocytes). Following the analysis of real-time PCR results, a significant down-regulation of *CDKAL1* expression was detected in activated and proliferating lymphocytes, compared to non-activated, resting cells (Fig. 4.3.3A). In order to exclude the possibility that mitogen activation had resulted in the differential expression of the housekeeping gene (*PPIA*) used in the real-time PCR experiments, an additional analysis was performed using an alternative endogenous control (*RPLPO*). This confirmed the previous set of results, indicating that *CDKAL1* is genuinely down-regulated in activated T and B lymphocytes (Fig.4.3.3B).

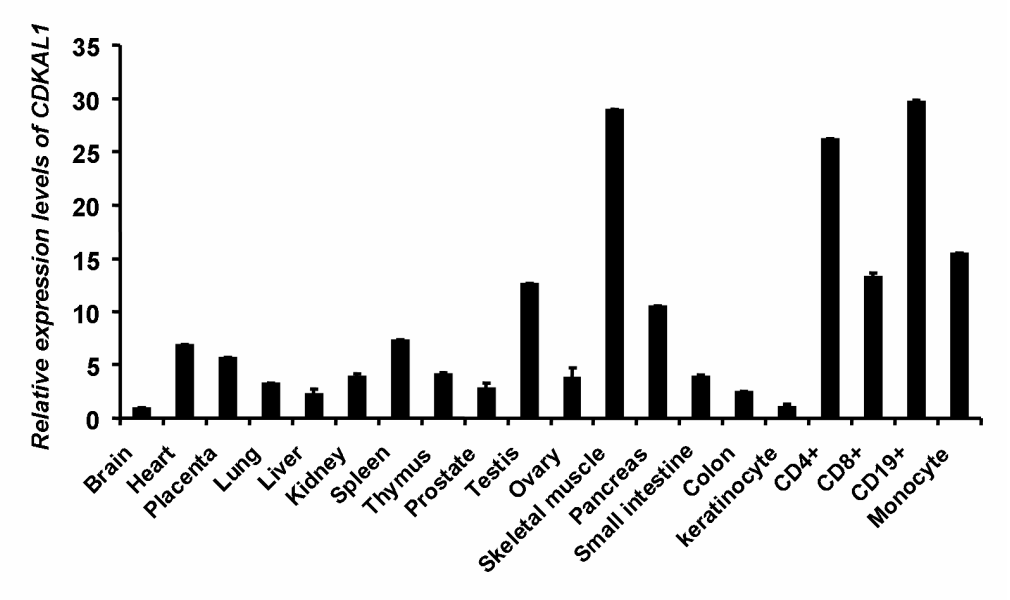


Figure 4.3.1: Real-time PCR analysis of *CDKAL1* gene expression

Transcript levels were quantified using the *PPIA* housekeeping gene as an endogenous control. The *CDKAL1/PPIA* ratio in brain was set as a baseline value to which all transcript levels were compared. $CD4^+$: $CD4^+$ T lymphocytes; $CD8^+$: $CD8^+$ T lymphocytes; $CD19^+$: $CD19^+$ B lymphocytes. Error bars refer to technical duplicates in real-time PCR experiments.

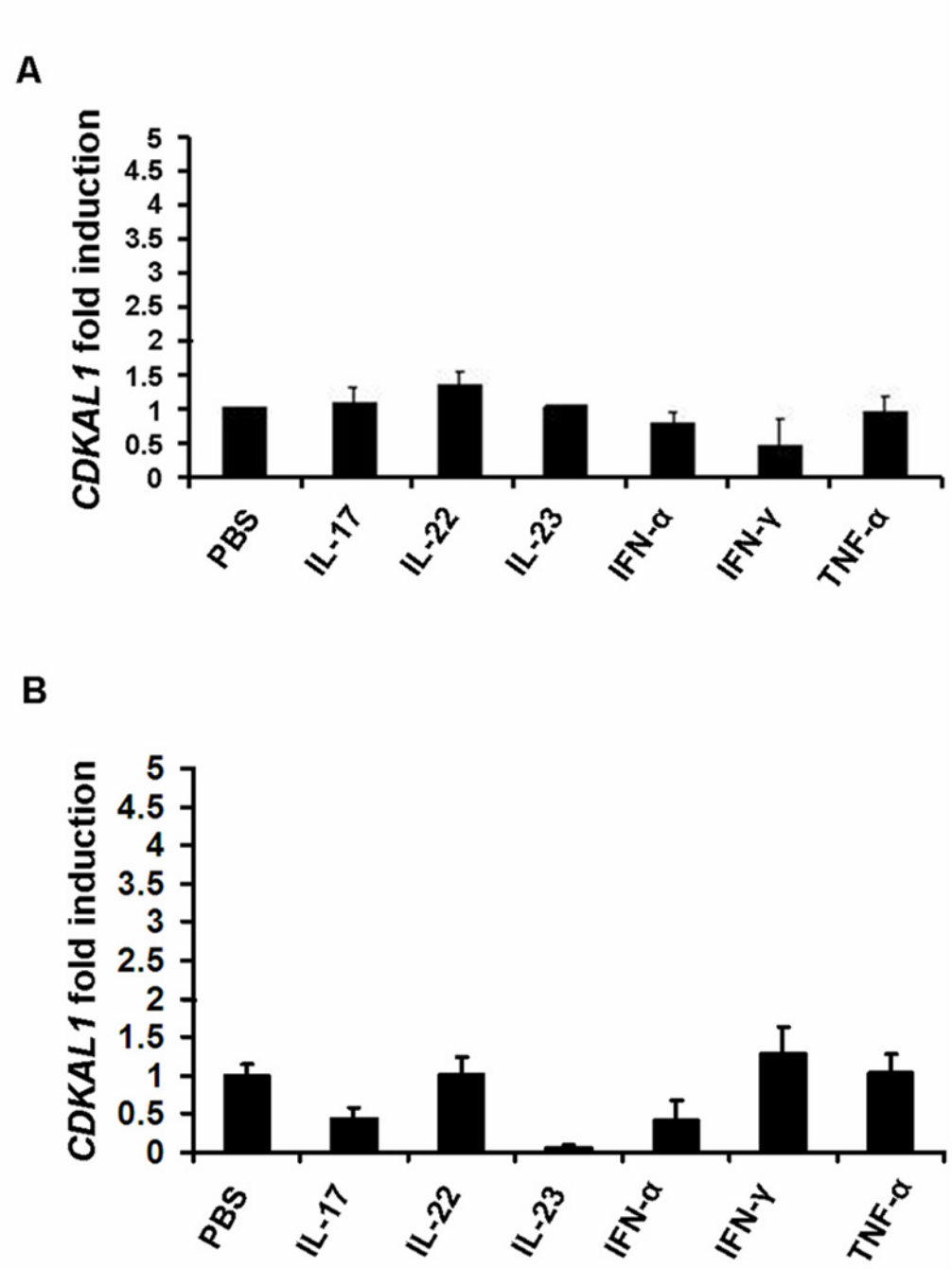

A
CDKAL1 fold induction
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
PBS
IL-17
IL-22
IL-23
IFN-α
IFN-γ
TNF-α
B
CDKAL1 fold induction
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
PBS
IL-17
IL-22
IL-23
IFN-α
IFN-γ
TNF-α

Figure 4.3.2: **Real-time PCR analysis of *CDKAL1* expression in primary keratinocytes**

Following cytokine stimulation, (100ng/mL for 24 hours for IL-17, IL-22, IL-23, IFN-γ and TNF-α; 5 hours for IFN-α), no differences in *CDKAL1* expression were detected in normal keratinocytes (**A**). Conversely, a marked *CDKAL1* down-regulation was detected in psoriatic keratinocytes treated with IL-17, IL-23 and IFN-α (**B**). The *PPIA* housekeeping gene was used as an endogenous control and the *CDKAL1*/*PPIA* ratio for PBS treated samples was set as a baseline value to which all transcript levels were normalized. Error bars refer to technical duplicates in real-time PCR experiments. The dashed horizontal line indicates the threshold of two-fold up-regulation of gene expression.

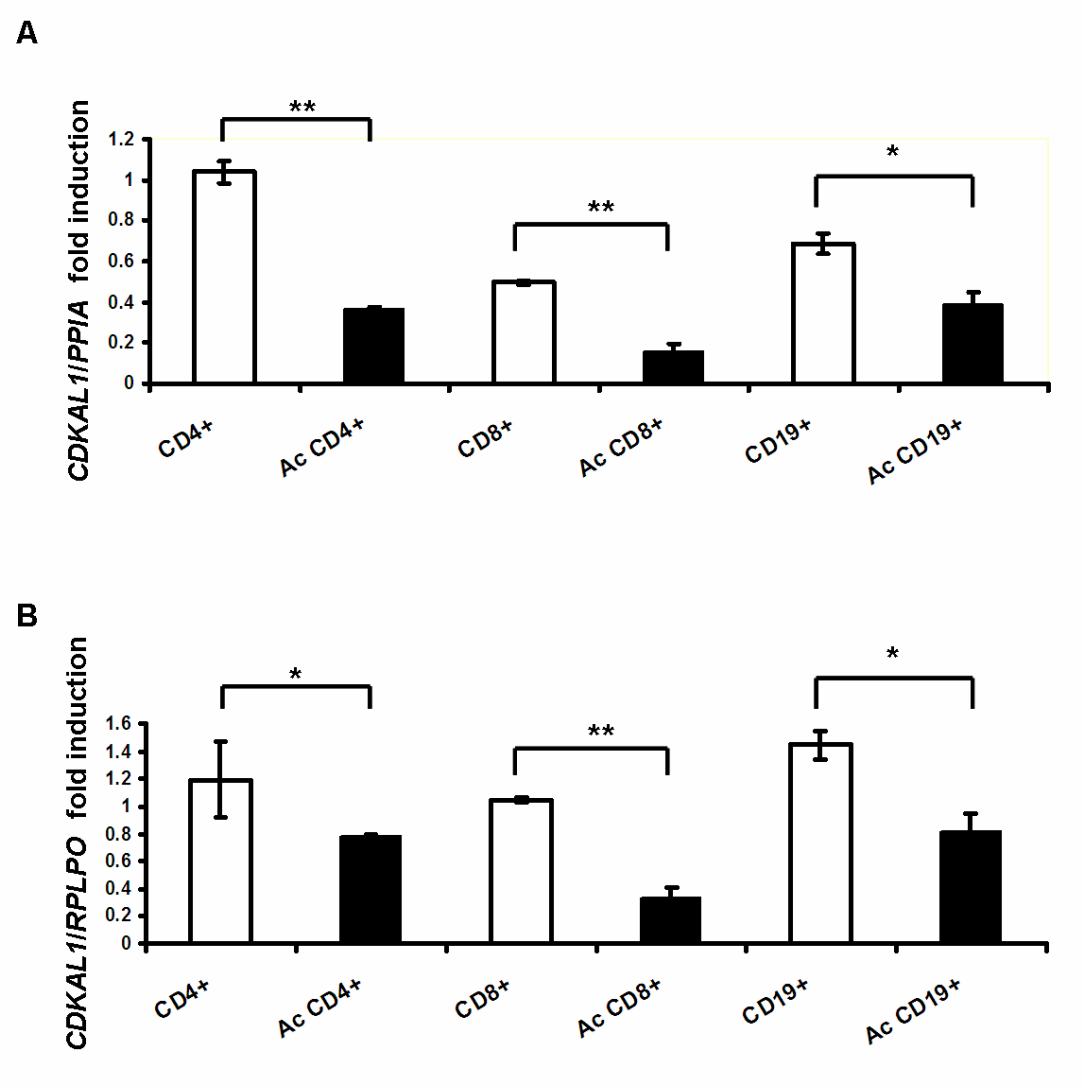


Figure 4.3.3: Real-time PCR analysis of *CDKAL1* expression in resting and activated lymphocytes

A significant *CDKAL1* down-regulation was detected in activated T and B cells. The *CDKAL1/PPIA* ratio for $CD4^+$ T lymphocytes was set as a baseline value, to which all transcript levels were normalized (**A**). The *CDKAL1/RPLPO* ratio for $CD4^+$ T lymphocytes was set as a baseline value (**B**). Error bars refer to technical duplicates in real-time PCR experiments. $CD4^+$: resting $CD4^+$ T lymphocytes; Ac $CD4^+$: activated $CD4^+$ T lymphocytes; $CD8^+$: resting $CD8^+$ T lymphocytes; Ac $CD8^+$: activated $CD8^+$ T lymphocytes; $CD19^+$: resting $CD19^+$ B lymphocytes. Ac $CD19^+$: activated $CD19^+$ B lymphocytes. Significance values were calculated using the t-test. *$p< 0.05$; **$p< 0.01$.

4.4 Discussion

The aim of the experiments described in this chapter was to explore *CDKAL1* expression, with a view to obtaining some preliminary insights into the mechanisms through which the gene may contribute to disease susceptibility. The first phase of the study consisted of a bioinformatic analysis of *CDKAL1* transcripts, which revealed the presence of five distinct mRNA isoforms. Importantly, one of these transcripts (*CDKAL1*-202*)* lacked two coding exons and was therefore expected to generate a smaller protein product than the one encoded by the full-length mRNA. For this reason, the tissue distribution of isoform *CDKAL1*-202 was investigated at the experimental level, by analyzing a panel of cDNAs derived from 11 tissues and cell types. These experiments failed to detect the presence of the *CDKAL1*-202 transcript in any of the examined tissues. At the same time, the annotation data available in the Ensembl genome browser indicate that *CDKAL1*-202 is expressed in at least one of the analysed tissues (placenta, which is the source of the EST sequence supporting the existence of *CDKAL1*-202). Thus it is still possible that *CDKAL1*-202 may be present in the examined samples, but at levels that are too low for detection by conventional RT-PCR.

All *CDKAL1* isoforms differ from the full length transcript in the length of their 5' (*CDKAL1*-001) and 3'UTRs (*CDKAL1*-001*, CDKAL1*-202*, CDKAL1*-203). Since 5' UTRs contain ribosome binding sites (Pickering and Willis 2005) and 3'UTRs are involved in the stabilization of mRNAs (Muhlrad and Parker 1999), it is possible that variations in the length of these regions may have an effect on gene expression levels, resulting in the differential regulation of the various isoforms.

Limited information is available on *CDKAL1* tissue distribution, as published studies mainly focused on tissues relevant to the pathogenesis of T2D (Scott, Mohlke et al. 2007; Zeggini, Weedon et al. 2007). In order to generate a broader overview of

CDKAL1 tissue distribution, a comprehensive characterization of transcript levels was performed, using a panel of cDNAs derived from 20 cell types. In agreement with published data (Scott, Mohlke et al. 2007), this experiment highlighted a very strong *CDKAL1* expression in skeletal muscle, a tissue that plays an important role in the pathogenesis of T2D. Of note, RT-PCR analyses also revealed abundant *CDKAL1* expression in several immune cell types, suggesting that *CDKAL1* might have a role in the inflammatory responses that underlie psoriasis and CD susceptibility. This observation was further supported by the analysis of cDNAs obtained from resting and activated T and B cells, which showed a marked change in *CDKAL1* expression, upon lymphocyte activation.

RT-PCR experiments also revealed modest *CDKAL1* levels in primary keratinocytes. Although these cells had only been passaged for a limited number of times, it is still possible that the culture conditions may have influenced the expression of *CDKAL1*. Thus, it is difficult to compare the transcript levels observed in these cells, with those detected in fresh tissues. Of note, *CDKAL1* down-regulation was observed in primary psoriatic keratinocytes after treatment with IL-17, IL-23 and IFN-α, three pro-inflammatory cytokines that have emerged as key players in the immunopathogenesis of psoriasis (Nestle and Gilliet 2005; Di Cesare, Di Meglio et al. 2009). These expression data were generated in keratinocytes obtained from a single patient and should therefore be interpreted with caution. However, it is noteworthy that the results of these experiments further support the idea that *CDKAL1* may be involved in inflammatory processes.

Taken together, the findings presented in this chapter indicate that *CDKAL1* is a widely expressed gene which might be differentially regulated by a variety of *cis*- (UTRs of different length) and *trans*- (cytokines) acting factors. The expression data also suggest

that *CDKAL1* might contribute to disease pathogenesis by affecting inflammatory responses.

Chapter 5: Functional analysis of the CDKAL1 protein

5.1 Bioinformatics

The homology data annotated in the Ensembl genome browser indicate that *CDKAL1* may be a member of the methylthiotransferase (MTTase) gene family (Fig 5.1.1), which encodes several proteins promoting the methylthiolation of tRNAs. This process, which is essential for accurate protein translation by the ribosome, involves the modification of the adenosine adjacent to the 3'end of the anticodon (Arragain, Handelman et al. 2010). MTTases typically include three functional domains: an N-terminal MTTase motif, a central radical-*S*-adenosylmethionine (SAM) and a C-terminal TRM2 and MiaB (TRAM) domain (Anton, Saleh et al. 2008), all of which appear to be present within the CDKAL1 protein sequence (Fig 5.1.1). Comparative analyses show that the portions of CDKAL1 corresponding to these three putative domains are very strongly conserved (Fig 5.1.2). The functional importance of these regions is also consistent with the lack of non-conservative SNP within the corresponding CDKAL1 exons (Fig. 5.1.3).

The Ensembl annotation for the CDKAL1 protein also indicates the presence of a putative transmembrane motif (Fig. 5.1.1). However, the poor evolutionary conservation of this region (Fig. 5.1.2) and the presence of deleterious sequence variants (Fig. 5.1.3) suggest that this prediction should be treated with great caution.

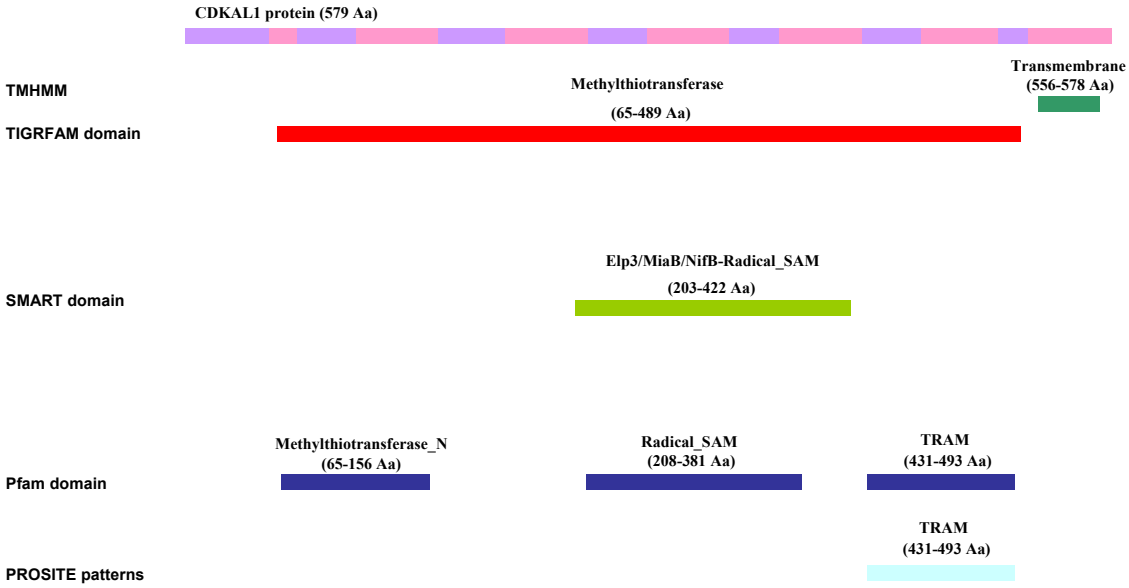


Figure 5.1.1: Graphic representation of CDKAL1 domains

Coloured bars indicate the position of the domains annotated in the databases listed on the left. The bar on the top indicates the position of *CDKAL1* exons. Aa: amino acid.

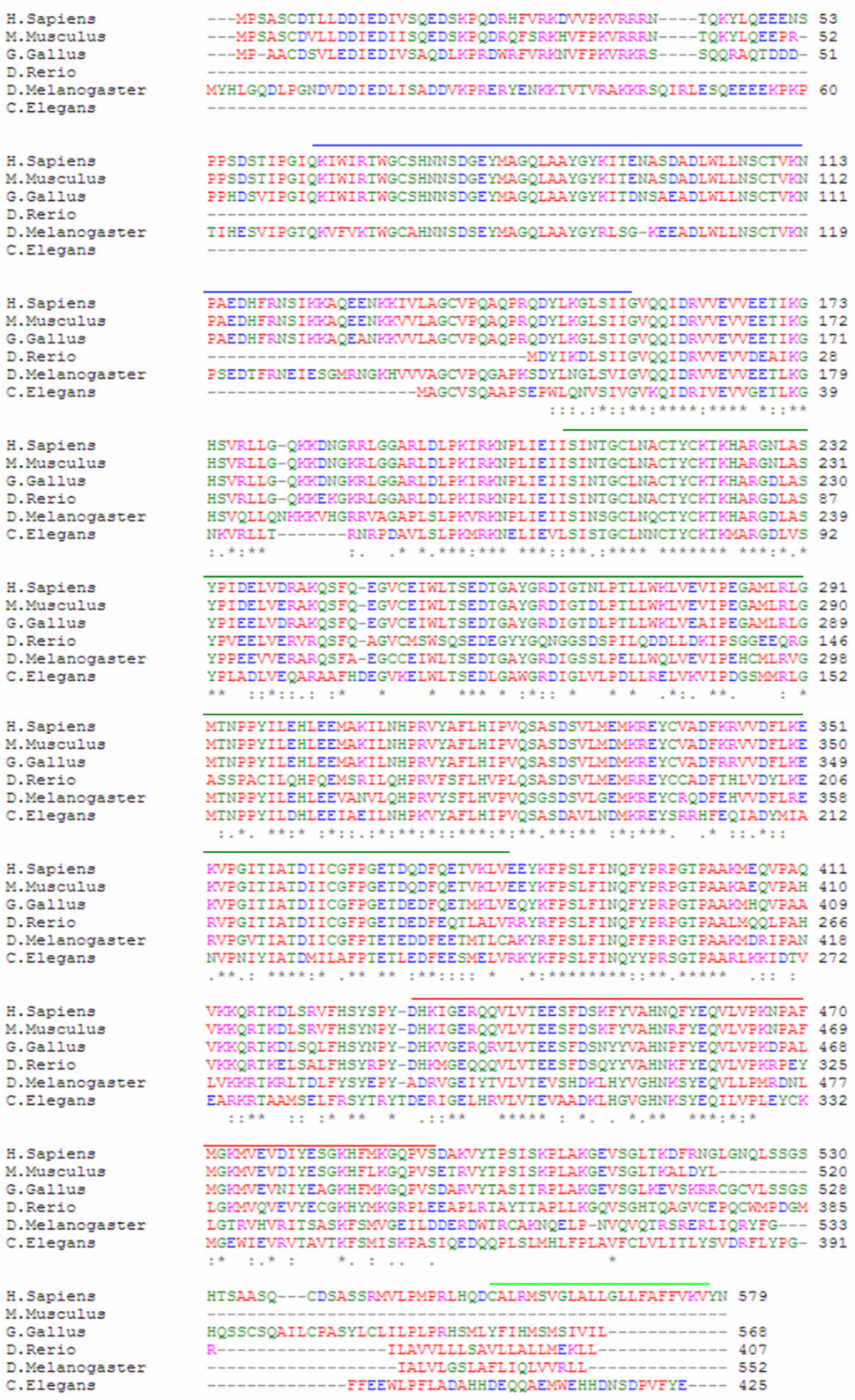

H.Sapiens ---MPSASCDTLLDDIEDIVSQEDSKPQDRHFVRKDVVPKVRRRN----TQKYLQEEENS 53
M.Musculus ---MPSASCDVLLDDIEDIISQEDSKPQDRQFSRKHVFPKVRRRN----TQKYLQEEPR- 52
G.Gallus ---MP-AACDSVLEDIEDIVSAQDLKPRDWRFVRKNVFPKVRKRS----SQQRAQTDDD- 51
D.Rerio ----------
D.Melanogaster MYHLGQDLPGNDVDDIEDLISADDVKPRERYENKKTVTVRAKKRSQIRLESQEEEEKPKP 60
C.Elegans ----------
H.Sapiens PPSDSTIPGIQKIWIRTWGCSHNNSDGEYMAGQLAAYGYKITENASDADLWLLNSCTVKN 113
M.Musculus PPSDSTIPGIQKIWIRTWGCSHNNSDGEYMAGQLAAYGYKITENASDADLWLLNSCTVKN 112
G.Gallus PPHDSVIPGIQKIWIRTWGCSHNNSDGEYMAGQLAAYGYKITDNSAEADLWLLNSCTVKN 111
D.Rerio ----------
D.Melanogaster TIHESVIPGTQKVFVKTWGCAHNNSDSEYMAGQLAAYGYRLSG-KEEADLWLLNSCTVKN 119
C.Elegans ----------
H.Sapiens PAEDHFRNSIKKAQEENKKIVLAGCVPQAQPRQDYLKGLSIIGVQQIDRVVEVVEETIKG 173
M.Musculus PAEDHFRNSIKKAQEENKKVVLAGCVPQAQPRQDYLKGLSIIGVQQIDRVVEVVEETIKG 172
G.Gallus PAEDHFRNSIKKAQEANKKVVLAGCVPQAQPRQDYLKGLSIIGVQQIDRVVEVVEETIKG 171
D.Rerio ----------MDYIKDLSIIGVQQIDRVVEVVDEAIKG 28
D.Melanogaster PSEDIFRNEIESGMRNGKHVVVAGCVPQGAPKSDYLNGLSVIGVQQIDRVVEVVEETLKG 179
C.Elegans ----------MAGCVSQAAPSEPWLQNVSIVGVKQIDRIVEVVGETLKG 39
:::.:*::**:****:**** *::**
H.Sapiens HSVRLLG-QKKDNGRRLGGARLDLPKIRKNPLIEIISINTGCLNACTYCKTKHARGNLAS 232
M.Musculus HSVRLLG-QKKDNGKRLGGARLDLPKIRKNPLIEIISINTGCLNACTYCKTKHARGNLAS 231
G.Gallus HSVRLLG-QKKDNGKRLGGARLDLPKIRKNPLIEIISINTGCLNACTYCKTKHARGDLAS 230
D.Rerio HSVRLLG-QKKEKGKRLGGARLDLPKIRKNPLIEIISINTGCLNACTYCKTKHARGDLAS 87
D.Melanogaster HSVQLLQNKKKVHGRRVAGAPLSLPKVRKNPLIEIISINSGCLNQCTYCKTKHARGDLAS 239
C.Elegans NKVRLLT-------RNRPDAVLSLPKMRKNELIEVLSISTGCLNNCTYCKTKMARGDLVS 92
:.*:** :. .* *.***:*** ***::**.:**** ******* ***:*.*
H.Sapiens YPIDELVDRAKQSFQ-EGVCEIWLTSEDTGAYGRDIGTNLPTLLWKLVEVIPEGAMLRLG 291
M.Musculus YPIDELVERAKQSFQ-EGVCEIWLTSEDTGAYGRDIGTDLPTLLWKLVEVIPEGAMLRLG 290
G.Gallus YPIEELVDRAKQSFQ-EGVCEIWLTSEDTGAYGRDIGTDLPTLLWKLVEAIPEGAMLRLG 289
D.Rerio YPVEELVERVRQSFQ-AGVCMSWSQSEDEGYYGQNGGSDSPILQDDLLDKIPSGGEEQRG 146
D.Melanogaster YPPEEVVERARQSFA-EGCCEIWLTSEDTGAYGRDIGSSLPELLWQLVEVIPEHCMLRVG 298
C.Elegans YPLADLVEQARAAFHDEGVKELWLTSEDLGAWGRDIGLVLPDLLRELVKVIPDGSMMRLG 152
** ::*::.: :* * * *** * :*:: * * * .*:. **. : *
H.Sapiens MTNPPYILEHLEEMAKILNHPRVYAFLHIPVQSASDSVLMEMKREYCVADFKRVVDFLKE 351
M.Musculus MTNPPYILEHLEEMAKILNHPRVYAFLHIPVQSASDSVLMDMKREYCVADFKRVVDFLKE 350
G.Gallus MTNPPYILEHLEEMAKILNHPRVYAFLHIPVQSASDSVLMDMKREYCVADFRRVVDFLKE 349
D.Rerio ASSPACILQHPQEMSRILQHPRVFSFLHVPLQSASDSVLMEMRREYCCADFTHLVDYLKE 206
D.Melanogaster MTNPPYILEHLEEVANVLQHPRVYSFLHVPVQSGSDSVLGEMKREYCRQDFEHVVDFLRE 358
C.Elegans MTNPPYILDHLEEIAEILNHPKVYAFLHIPVQSASDAVLNDMKREYSRRHFEQIADYMIA 212
:.*. **:* :*::.:*:**:*::***:*:**.**:** :*:***. .* ::.*::
H.Sapiens KVPGITIATDIICGFPGETDQDFQETVKLVEEYKFPSLFINQFYPRPGTPAAKMEQVPAQ 411
M.Musculus KVPGITIATDIICGFPGETDQDFQETVKLVEEYKFPSLFINQFYPRPGTPAAKAEQVPAH 410
G.Gallus KVPGITIATDIICGFPGETDEDFQETMKLVEQYKFPSLFINQFYPRPGTPAAKMHQVPAA 409
D.Rerio RVPGITIATDIICGFPGETDEDFEQTLALVRRYRFPSLFINQFYPRPGTPAALMQQLPAH 266
D.Melanogaster RVPGVTIATDIICGFPTETEDDFEETMTLCAKYRFPSLFINQFFPRPGTPAAKMDRIPAN 418
C.Elegans NVPNIYIATDMILAFPTETLEDFEESMELVRKYKFPSLFINQYYPRSGTPAARLKKIDTV 272
.**.: ****:* .** ** :**:::: * .*:********::**.***** .:: :
H.Sapiens VKKQRTKDLSRVFHSYSPY-DHKIGERQQVLVTEESFDSKFYVAHNQFYEQVLVPKNPAF 470
M.Musculus VKKQRTKDLSRVFHSYNPY-DHKIGERQQVLVTEESFDSKFYVAHNRFYEQVLVPKNPAF 469
G.Gallus VKKQRTKDLSQLFHSYNPY-DHKVGERQRVLVTEESFDSNYYVAHNPFYEQVLVPKDPAL 468
D.Rerio VKKQRTKELSALFHSYRPY-DHKMGEQQQVLVTEESFDSQYYVAHNKFYEQVLVPKRPEY 325
D.Melanogaster LVKKRTKRLTDLFYSYEPY-ADRVGEIYTVLVTEVSHDKLHYVGHNKSYEQVLLPMRDNL 477
C.Elegans EARKRTAAMSELFRSYTRYTDERIGELHRVLVTEVAADKLHGVGHNKSYEQILVPLEYCK 332
::** :: :* ** * .::** ***** : *. . *.** ***:*:*
H.Sapiens MGKMVEVDIYESGKHFMKGQPVSDAKVYTPSISKPLAKGEVSGLTKDFRNGLGNQLSSGS 530
M.Musculus MGKMVEVDIYESGKHFLKGQPVSETRVYTPSISKPLAKGEVSGLTKALDYL---------- 520
G.Gallus MGKMVEVNIYEAGKHFMKGQPVSDARVYTASITRPLAKGEVSGLKEVSKRRCGCVLSSGS 528
D.Rerio LGKMVQVEVYECGKHYMKGRPLEEAPLRTAYTTAPLLKGQVSGHTQAGVCEPQCWMPDGM 385
D.Melanogaster LGTRVHVRITSASKFSMVGEILDDERDWTRCAKNQELP-NVQVQTRSRERLIQRYFG--- 533
C.Elegans MGEWIEVRVTAVTKFSMISKPASIQEDQQPLSLMHLFPLAVFCLVLITLYSVDRFLYPG- 391
:* :.* : *. : .. . *
H.Sapiens HTSAASQ---CDSASSRMVLPMPRLHQDCALRMSVGLALLGLLFAFFVKVYN 579
M.Musculus ----------
G.Gallus HQSSCSQAILCPASYLCLILPLPRHSMLYFIHMSMSIVIL---------- 568
D.Rerio R----------ILAVVLLLSAVLLALLMEKLL---------- 407
D.Melanogaster ----------IALVLGSLAFLIQLVVRLL---------- 552
C.Elegans ----------FFEEWLPFLADAHHDEQQAEMWEHHDNSDPVFYE---- 425
Methylthiotransferase_N (65-156 Aa)
Radical_SAM (208-381 Aa)
TRAM (431-493 Aa)
Transmembrane (556-578 Aa)

Figure 5.1.2: Comparative analysis of the CDKAL1 protein sequence

The alignment of CDKAL1 protein sequences shows a high conservation of the MTTase, Radical_SAM and TRAM domains. Amino acids with similar chemical properties are highlighted with the same colour. "*": totally conserved residues; ":": conserved substitutions ".":semi-conserved substitutions. Aa: amino acid.

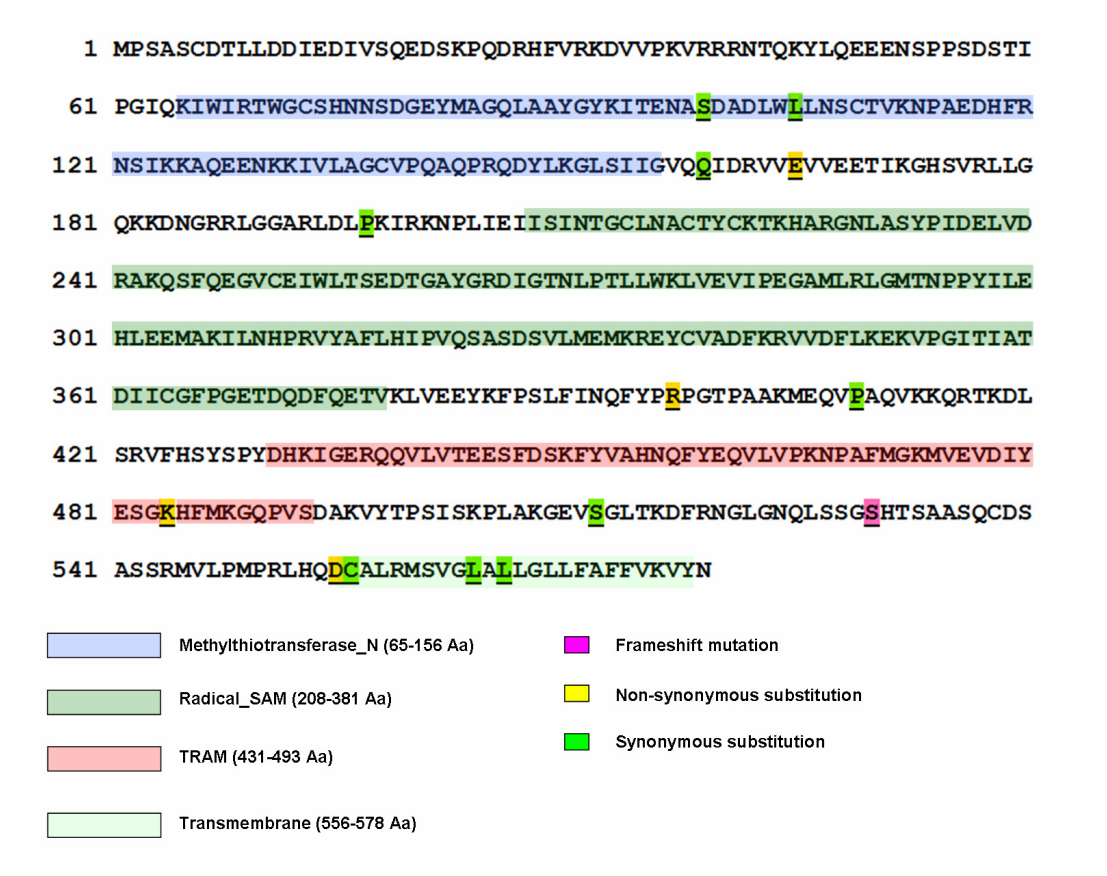


Figure 5.1.3: CDKAL1 amino acid sequence variation

The distribution of CDKAL1 sequence variants was assessed, based on the data reported in the Ensembl genome browser. This analysis showed that the predicted MTTase domains harbored only one amino acid substitution.

5.2 Validation of a α-CDKAL1 antibody

At the time these functional studies were first undertaken, only one α-CDKAL1 antibody was commercially available. When used in western blot experiments, the antibody detected a weak 65 kDa signal, consistent with the predicted size of the CDKAL1 protein, together with a stronger 55 kDa band, which was found in all examined cell types (Fig. 5.2.1). In order to elucidate which of the two bands, if any, corresponded to the CDKAL1 protein, a number of experiments were undertaken, with a view to assessing the antibody's specificity.

The ability of the antibody to recognize a recombinant CDKAL1 protein was first investigated. The *CDKAL1* full-length cDNA sequence was cloned into the pGEX-4T-1 vector (Fig. 5.2.2), which encodes a glutathione-S-transferase (GST) protein. A fusion protein was subsequently generated by transforming the GST-CDKAL1 construct in bacterial cells and inducing its expression by IPTG treatment. As shown in Figure 5.2.3, a time-dependent increase in the intensity of a 91 kDa band (consistent with the combined molecular weight of the GST moiety and a 65 kDa protein) was observed in the bacterial cultures, following IPTG-treatment. The pre- and post-induction samples were next analysed by western blot, using α-CDKAL1 and α-GST antibodies. As shown in Figure 5.2.4, both antibodies detected a clear signal at 91 kDa, demonstrating that the protein induced by IPTG treatment and recognized by α-CDKAL1 is indeed GST-CDKAL1.

The specificity of the antibody was next validated in cells where CDKAL1 expression had been transiently knocked-down by RNA interference (RNAi). A specific down-regulation of the 65 kDa band was observed in immortalized keratinocytes (HaCaT cell line) (Fig. 5.2.5), whereas the intensity of the 55 kDa signal remained unchanged. These data further support the notion that the 65 kDa band is generated by binding of the

antibody to CDKAL1.

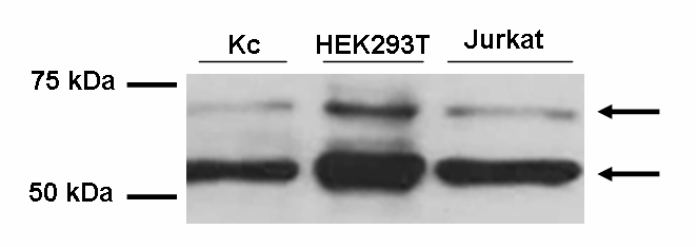


Figure 5.2.1: CDKAL1 western blot analysis

The expression of a 55 kDa and a 65 kDa band (highlighted by black arrows) was detected in all examined cell types. Kc: cultured primary keratinocyes.

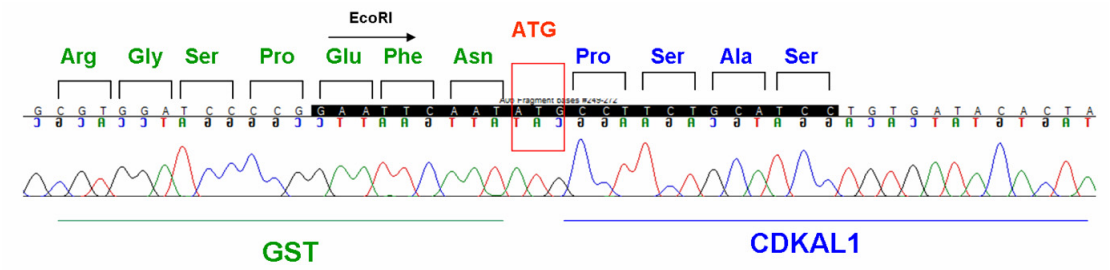


Figure 5.2.2: GST-CDKAL1 sequence analysis

Plasmid sequence analysis confirmed that the *CDKAL1* cDNA had been cloned in frame with the GST moiety. Green and blue fonts respectively indicate the GST and CDKAL1 amino acidic sequences. The red box shows the position of *CDKAL1* start codon. The black arrow indicates the position of the EcoRI cloning site.

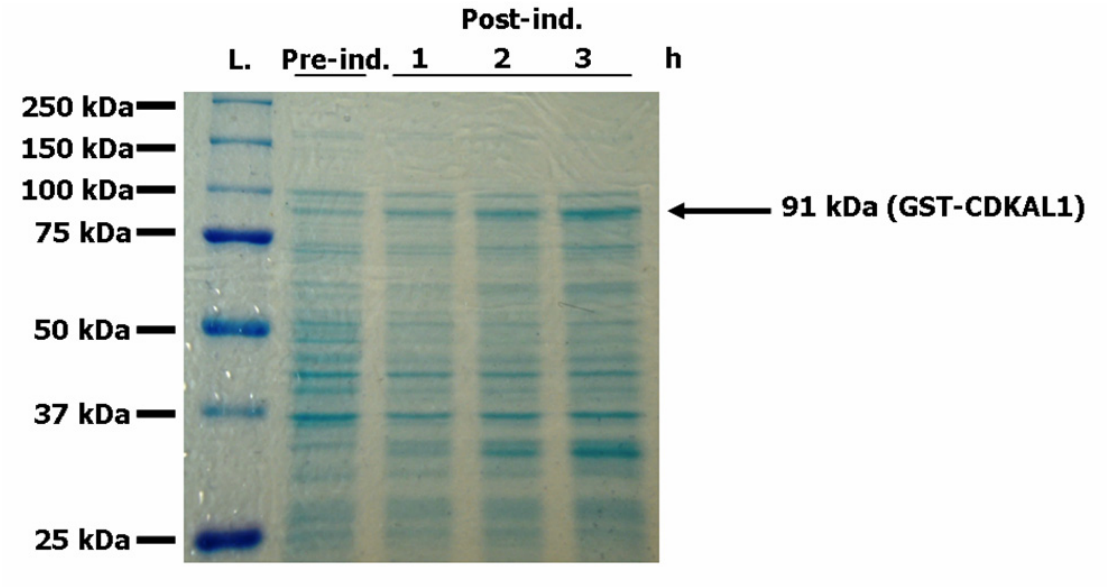


Figure 5.2.3: Induction of GST-CDKAL1 expression

A time-dependent increase of the GST-CDKAL1 band intensity was detected after isopropyl β-D-1-thiogalactopyranoside (IPTG) treatment. L: molecular weight ladder; pre-ind.: pre-induction sample; post-ind.: post-induction samples; h.: hours.

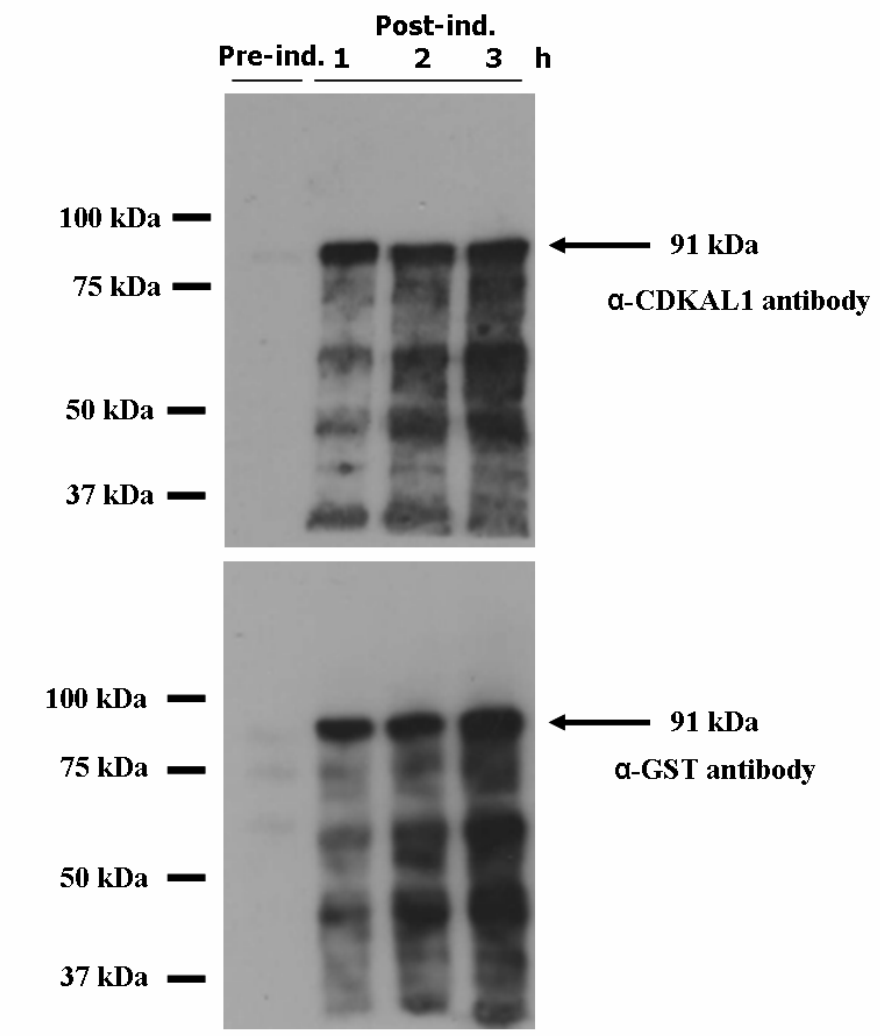


Figure 5.2.4: Western blot analysis of IPTG induction

The same 91 kDa band (corresponding to GST-CDKAL1) was detected using both α-CDKAL1 and α-GST antibodies. L: molecular weight ladder; pre-ind.: pre-induction sample; post-ind.: post-induction samples; h.: hours.

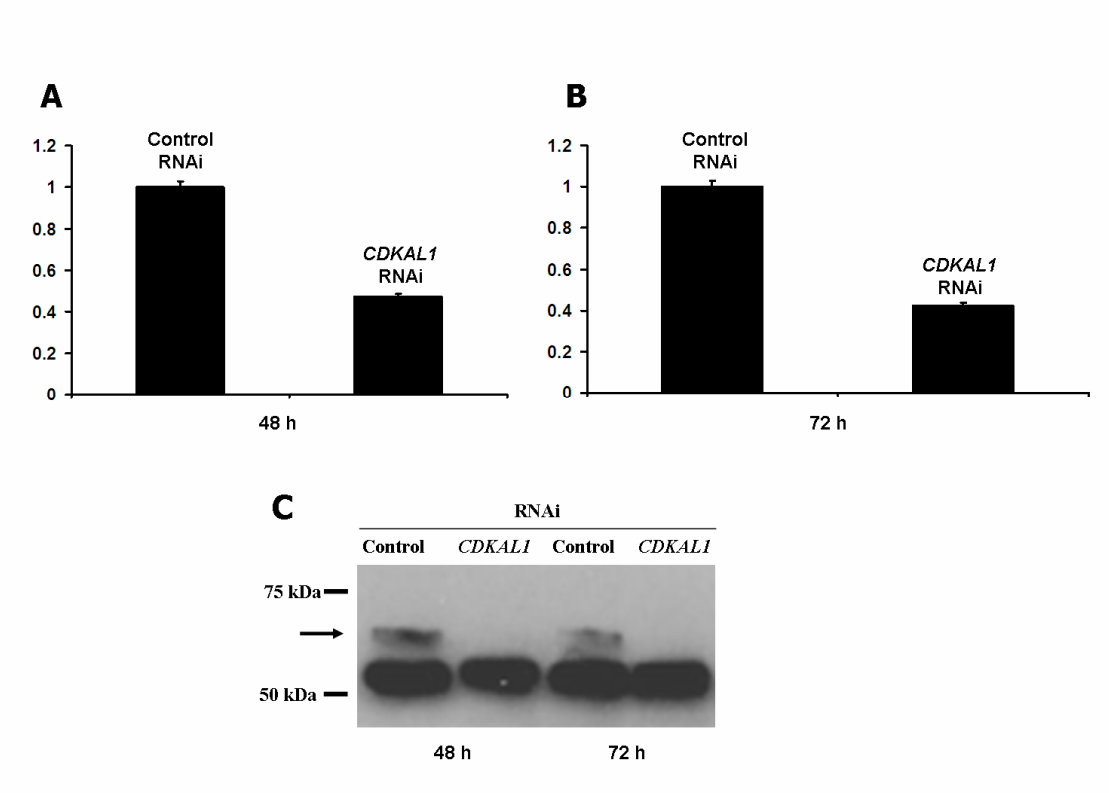


Figure 5.2.5: α-CDKAL1 antibody specificity characterization

RNAi treatment caused a marked down-regulation of *CDKAL1* mRNA (panels **A** and **B**). At the protein level (panel **C**), a specific down-regulation of the 65 kDa band (highlighted by a black arrow) was observed, whereas no change in intensity was detected for the 55 kDa signal.

5.3 CDKAL1 subcellular localization

Following its validation, the α-CDKAL1 antibody was used in western blot and immunofluorescence experiments, with the aim of establishing CDKAL1 subcellular localization. Western blots of extracts obtained from HaCaT and Jurkat cell lines demonstrated that CDKAL1 expression can only be detected in the cytoplasmic fraction of these cells (Fig. 5.3.1). The analysis of known nuclear (lamin-α) and cytoplasmic (β-actin) proteins confirmed the purity of the two fractions.

The localization of CDKAL1 was also verified by IF staining of HaCaT cells. This showed a strong cytoplasmic staining for CDKAL1, with no signal detected in the nucleus (Fig. 5.3.2).

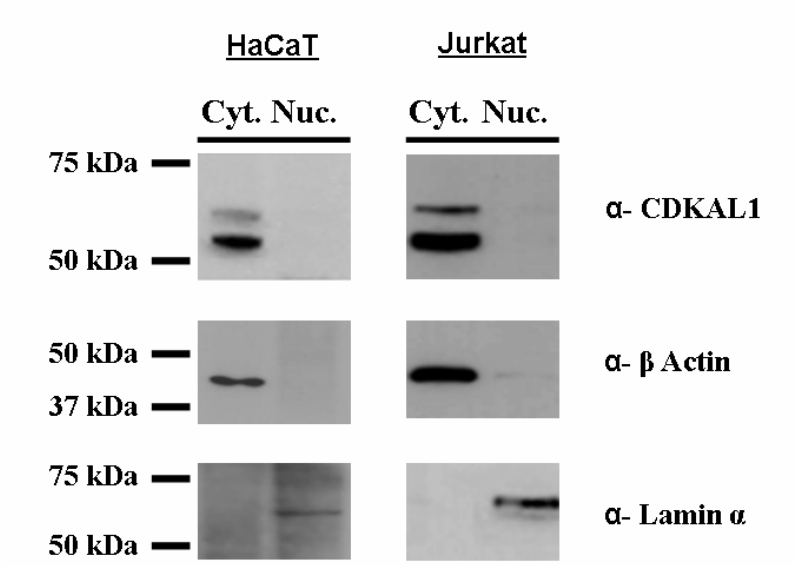


Figure 5.3.1: CDKAL1 is expressed in the cytoplasmic cell fraction

CDKAL1 expression is only detectable in the cytoplasmic fraction of HaCaT and Jurkat cells. Detection of known nuclear (lamin-α) and cytoplasmic (β-actin) proteins confirmed the purity of the nuclear (Nuc.) and cytoplasmic (Cyt.) fractions.

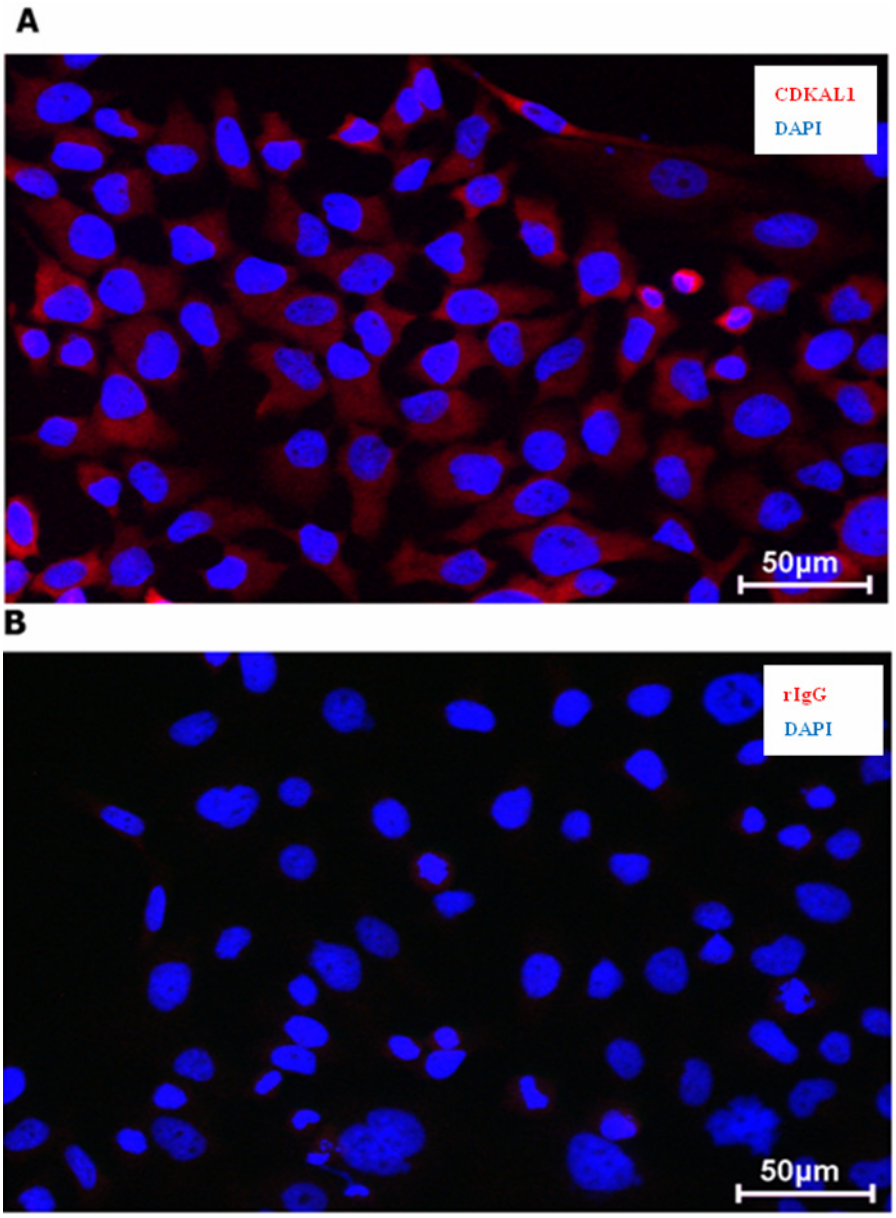


Figure 5.3.2: CDKAL1 localizes to the cell cytoplasm

IF staining of HaCaT cells confirmed the cytoplasmic localization of CDKAL1 (**A**), while cells stained with an isotype control antibody showed no signal (**B**). Nuclei were stained in blue with DAPI.

5.4 Generation of stable *CDKAL1* knock-down cell lines

Stable *CDKAL1* knock-down cell lines were generated at the outset of the study as a tool to investigate the function of the CDKAL1 protein.

Two commercial vectors, each encoding a distinct short hairpin RNA (sh-A11 and sh-H6, see Table 5.4.1), were used to knock-down *CDKAL1* expression, in both HaCaT and Jurkat cell lines. A non-silencing plasmid, carrying a sh-RNA sequence with no homology to known mammalian genes (sh-non-sil.), was used as a negative control for both cell types. In all vectors, the sh-RNAs were encoded by a bicistronic transcript that also contained the turbo Green Fluorescent Protein (tGFP) gene.

Stable knock-down cell lines were generated by initially co-expressing the virus packaging vectors (pGIPZ-shRNA, VSV-G expressor and p8.91 gag-pol plasmids) in a HEK293T cell line. The efficiency of this procedure was demonstrated by the high number of tGFP-expressing cells (Fig. 5.4.1). The cell medium containing the viral particles was next used to transduce both HaCaT and Jurkat cells. The HaCaT transduction efficiency of cells was confirmed by the high frequency of tGFP-expressing cells (Fig. 5.4.2). In Jurkat cells, the transduction efficiency was assessed by flow cytometry, as optical estimation of fluorescence is difficult to achieve in cell lines growing as a suspension. As shown in Figure 5.4.3, a transduction efficiency exceeding 75% was estimated in these cells, based on tGFP fluorescence.

The degree of *CDKAL1* knock-down was next assessed by real-time PCR. No significant differences in the levels of *CDKAL1* expression were observed in the knock-down cells, compared to the controls transduced with the non-silencing construct (Fig. 5.4.4). These results suggested variability in the knock-down efficiency of both HaCaT and Jurkat cell lines, possibly due to the number of sh-RNA copies integrated in their genomes. To circumvent this problem, a number of clonal lines were generated from

single cells, selected for high levels of tGFP-expression (Fig. 5.4.5). Twelve HaCaT and six Jurkat clones were thus successfully expanded (Table 5.4.2 and 5.4.3). Real-time PCR analysis confirmed a significant down-regulation of *CDKAL1* mRNA in seven HaCaT (three sh-A11 and four sh-H6) and two Jurkat (both sh-H6) clones. The efficiency of knock-down was further confirmed in HaCaT and Jurkat cells that had been cultured for four weeks, thus demonstrating the stability of the clones (Tables 5.4.2 and 5.4.3; Fig. 5.4.6).

Table 5.4.1: Short hairpin sequences used to knock-down *CDKAL1* expression

sh-RNA	Open biosystems identifier	Hairpin sequence*	Target region
sh-A11	V2HS-173734	TGCTGTTGACAGTGAGCGCGCCTCGC CAGGACTACCTTAATAGTGAAGCCAC AGATGTATTAAGGTAGTCCTGGCGAG GCTTGCCTACTGCCTCGGA	exon 6
sh-H6	V2HS-173733	TGCTGTTGACAGTGAGCGCGCTCAAG AGGAGAACAAGAAATAGTGAAGCCA CAGATGTATTTCTTGTTCTCCTCTTGA GCTTGCCTACTGCCTCGGA	exon 6

*The mir-30 backbone sequence is represented by black font; the sense and antisense sh-RNA sequences are highlighted in red and blue, respectively; the loop sequence is in green.

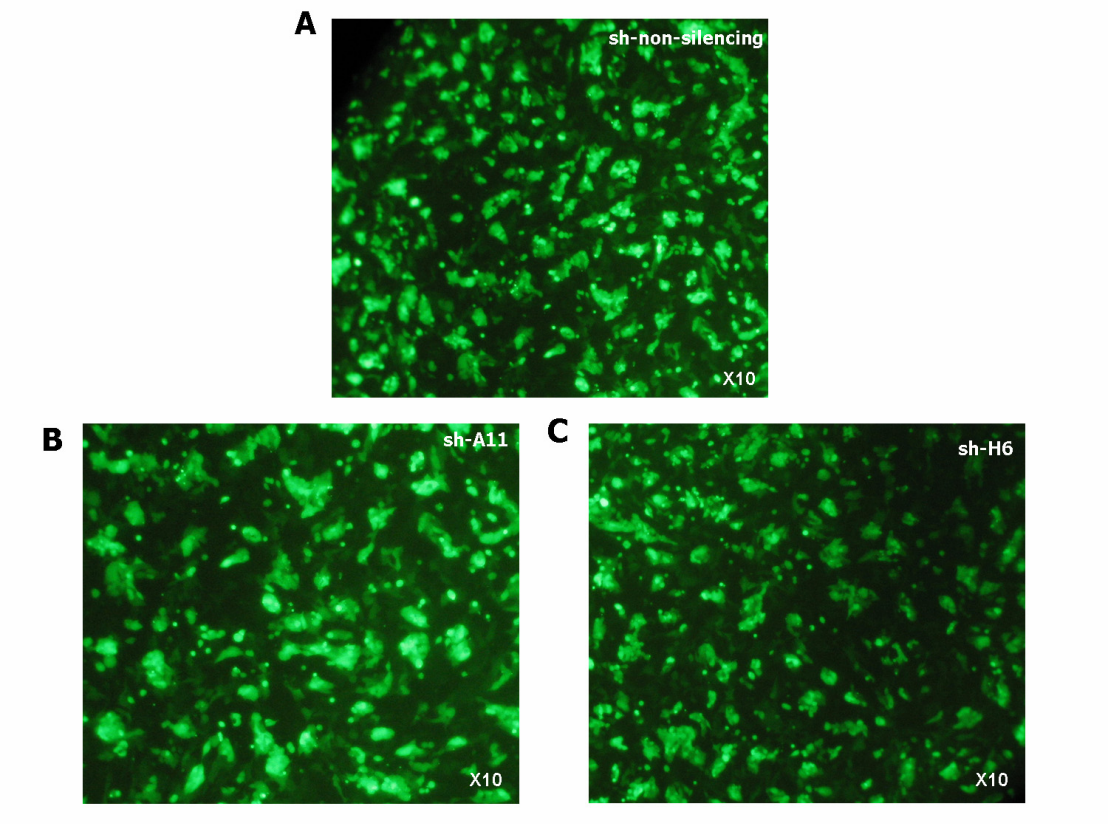


Figure 5.4.1: HEK293T cell line transfection efficiency

All three sh-RNA constructs were efficiently transfected into HEK293T cells, as demonstrated by the high number of tGFP expressing cells.

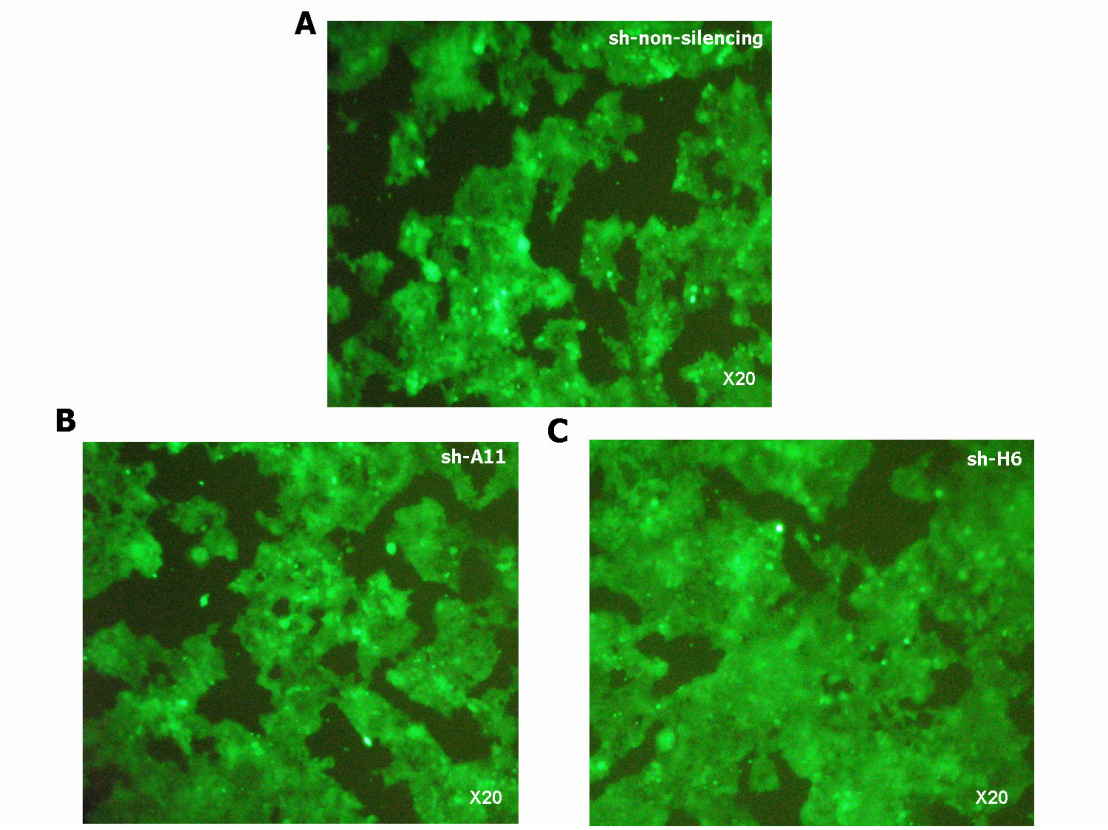


Figure 5.4.2: HaCaT cell line transduction efficiency

Lentiviral particles bearing the three sh-RNA constructs were efficiently transduced into HaCaT cells, as demonstrated by the high number of tGFP expressing cells.

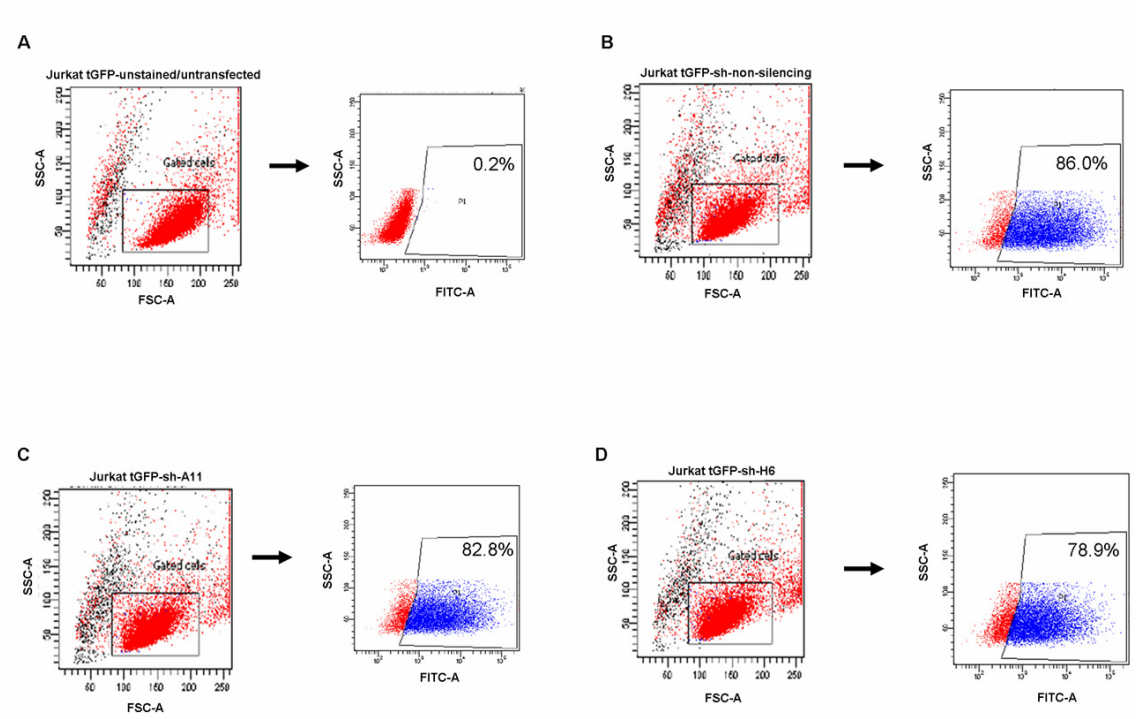


Figure 5.4.3: Jurkat cell line transduction efficiency

A transduction efficiency exceeding 75% was estimated for all transduced cells, based on tGFP staining. Cells were first gated based on their size and granularity, as measured by the Forward (FSC-A), and Side Scatter (SSC-A) (left-hand plots). Next, tGFP fluorescence was detected in the FITC-A channel and cells were gated, based on the fluorescence values observed in unstained and untransfected cells. The tGFP expressing cells thus detected are highlighted in blue. The percentage figures in the plots indicate the frequency of gated, tGFP expressing cells.

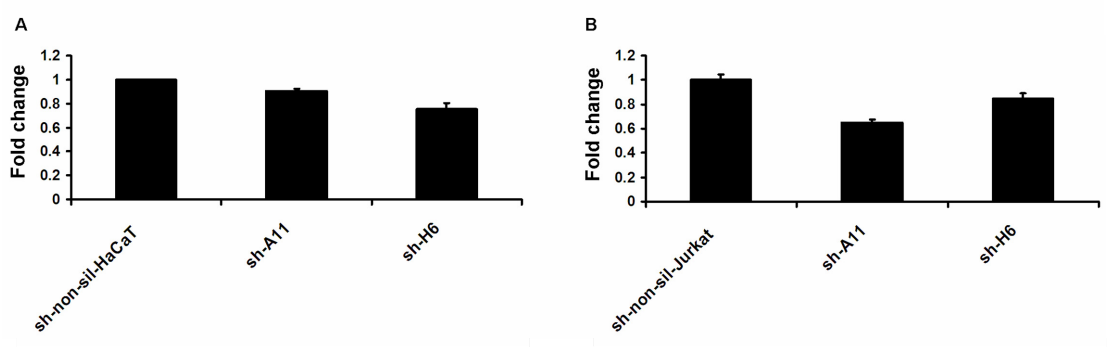


Figure 5.4.4: Real-time PCR analysis of *CDKAL1* gene expression in stably transduced HaCaT and Jurkat cell lines

No differences in *CDKAL1* expression were detected in cells transduced with sh-RNA silencing constructs, in either HaCaT (**A**) or and Jurkat cells (**B**). Transcript levels were quantified using the *PPIA* housekeeping gene as an endogenous control. Error bars refer to technical duplicates in real-time PCR experiments.

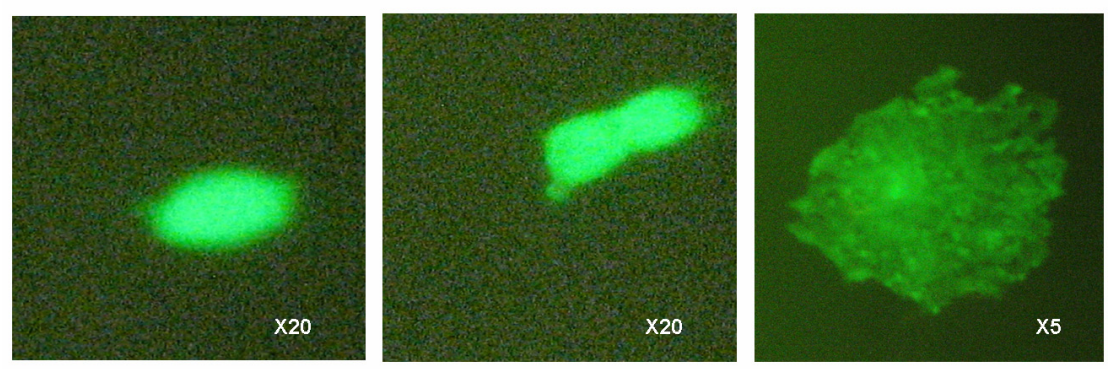


Figure 5.4.5: Clonal expansion of transduced HaCaT cells

Clones were obtained starting from a single, tGFP expressing cell (**left panel**), which divided through mitosis (**central panel**) until a colony was formed (**right panel**).

Table 5.4.2: sh-*CDKAL1* HaCaT clones

sh-RNA	clone name	*CDKAL1* down-regulation*
sh-non-silencing	sh-non-sil-HaCaT-1	-
	sh-non-sil-HaCaT-2	-
	sh-non-sil-HaCaT-3	-
	sh-non-sil-HaCaT-4	-
sh-A11	sh-A11-2	75%
	sh-A11-3	**70%**
	sh-A11-4	29%
	sh-A11-5	**72%**
sh-H6	sh-H6-1	79%
	sh-H6-2	**85%**
	sh-H6-3	**81%**
	sh-H6-4	69%

*Transcript abundance measured by real-time PCR and expressed as percentage of *CDKAL1* levels in silenced clones compared to non-silenced ones. The clones selected for further functional studies are shown in bold.

Table 5.4.3: sh-*CDKAL1* Jurkat clones

sh-RNA	**clone name**	***CDKAL1* down-regulation***
sh-non-silencing	sh-non-sil-Jurkat-1	-
	sh-non-sil-Jurkat-3	-
sh-A11	sh-A11-5-A	2%
	sh-A11-6-B	3%
sh-H6	**sh-H6-7-C**	**50%**
	sh-H6-8-D	**62%**

*Transcript abundance measured by real-time PCR and expressed as percentage of *CDKAL1* levels in silenced clones compared to the non-silenced ones. The clones selected for further functional studies are shown in bold.

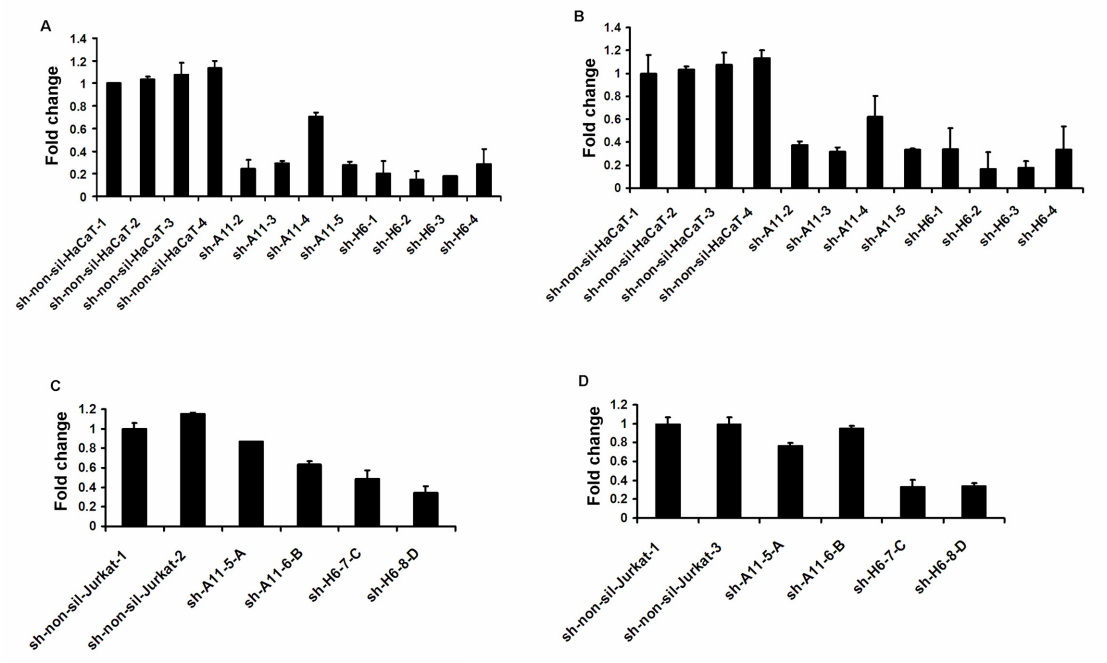


Figure 5.4.6: Real-time PCR analysis of *CDKAL1* gene expression in clones derived from stably transduced HaCaT and Jurkat cell lines

A significant down-regulation of *CDKAL1* expression was detected in the majority of analysed clones, in both HaCaT (**A**) and Jurkat cells (**C**). The knock-down efficiency was further confirmed in HaCaT (**B**) and Jurkat cells (**D**), after four weeks of culture. Transcript levels were quantified using the housekeeping *PPIA* gene as an endogenous control. Error bars refer to technical duplicates in real-time PCR experiments.

5.5 Cell cycle analysis of *CDKAL1* knock-down cells

The fact that *CDKAL1* lies in a chromosome region that is amplified in bladder tumours (Oeggerli, Schraml et al. 2006) suggests the possibility that the gene may be involved in the control of cell cycle progression. To explore this hypothesis, the effect of *CDKAL1* knock-down on cell proliferation was examined. In particular, the proliferation rates of HaCaT clones sh-A11-5 and sh-H6-2 were assessed at several time-points, using the WST-8 cell viability assay. A significant decrease in proliferation was observed in the knock-down cell lines, as early as one hour after the addition of the detection reagent (Fig. 5.5.1). This finding was further investigated by using flow cytometry to assess the DNA content of sh-A11-3 and sh-H6-3*CDKAL1* knock-down cells. As shown in Figure 5.5.2, an increase in the percentage of cells in G_0/G_1, accompanied by a decrease of those in S phase, was observed in the *CDKAL1* knock-down clones.

Taken together, these results suggest a possible involvement of *CDKAL1* in cell cycle progression.

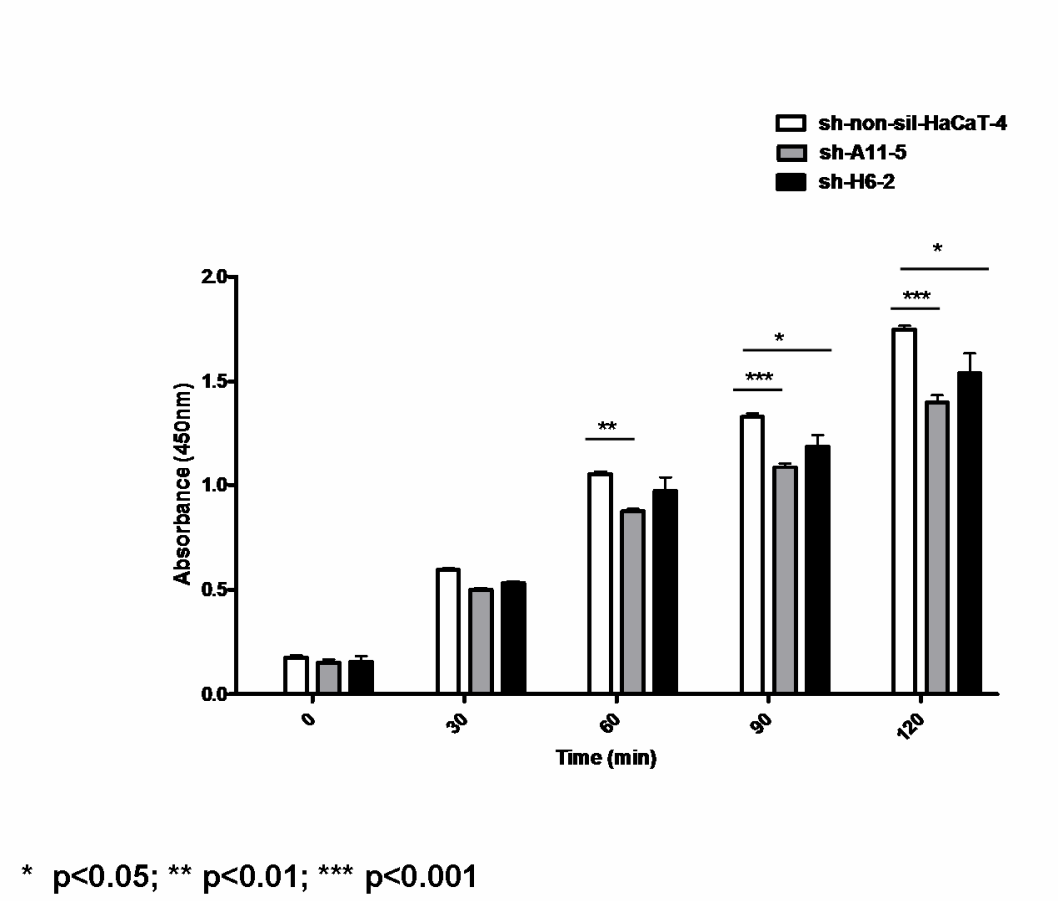
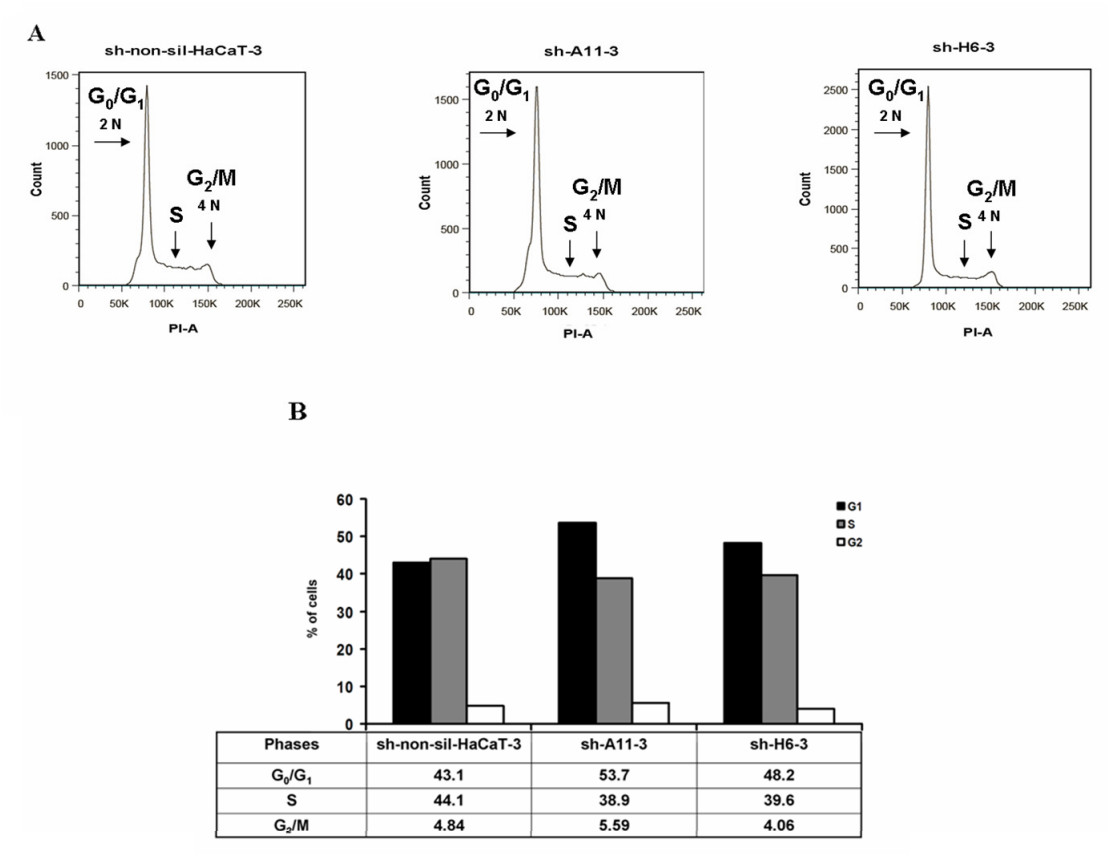


Figure 5.5.1: *CDKAL1* knock-down affects cell proliferation

CDKAL1 silencing results in a marked, time-dependent decrease in cell proliferation. Significance values were calculated using a two-way ANOVA test. Error bars refer to technical triplicates in WST-8 treatments.



Phases	sh-non-sil-HaCaT-3	sh-A11-3	sh-H6-3
G_0/G_1	43.1	53.7	48.2
S	44.1	38.9	39.6
G_2/M	4.84	5.59	4.06

Figure 5.5.2: *CDKAL1* knock-down influences the rate of cell cycle progression

Panel **A**: DNA content of stably transduced HaCaT cells. Gates were configured manually in order to determine the percentage of cells in G_0/G_1, S and G_2/M phase, based on DNA content. An increase in the percentage of cells in G_0/G_1, accompanied by a decrease of those in S phase, was observed in the *CDKAL1* knock-down clones. Panel **B**: the percentage of cells at G_0/G_1, S and G_2/M phases were graphically plotted.

5.6 Expression profiling of *CDKAL1* knock-down cells

Gene expression profiles were generated by comparing the transcriptome of HaCaT knock-down lines sh-A11-3, sh-A11-5, sh-H6-2 and sh-H6-3 with that of sh-non-sil-HaCaT-3 and sh-non-sil-HaCaT-4.

Two lists of differentially expressed genes were created by comparing: i) the two sh-A11 *vs.* the two HaCaT-non-sil. clones (822 transcripts; FDR ≤ 0.062), ii) the two sh-H6 *vs.* the two HaCaT-non-sil. clones (796 transcripts; FDR ≤ 0.01). The combined list included 1,342 genes potentially influenced by *CDKAL1* knock-down. In order to control for transcriptional variation between the two knock-down clones, subsequent analyses focused on the subset of genes (n = 276) that were differentially expressed in both cell lines. An examination of fold change log ratios showed that 274 out of the 276 differentially expressed genes followed the same trend in both knock-down clones (Fig. 5.6.1). Next, two lists, respectively including 118 up- and 156 down-regulated genes were generated, and a gene ontology analysis was performed, in order to identify the cellular pathways that were most affected by *CDKAL1* knock-down. This showed that the down-regulated genes were mainly involved in amino acid metabolism, cell adhesion, apoptosis and inflammatory processes. Conversely, the up-regulated genes were linked to pathways related to metabolism, transcription, cell adhesion, remodelling and cell cycle (Fig. 5.6.2).

In order to identify the genes that were most suitable for follow-up, three transcript sets were combined: i) the top 8 up-regulated genes, selected by the GeneGo software, based on their differential expression and their involvement in known cellular pathways; ii) the top 16 down-regulated genes, selected by the GeneGo software, based on their differential expression and their involvement in known cellular pathways; iii) a manually curated gene selection, including the ten transcripts associated with the most

significant p values and expression fold changes. The use of these selection criteria yielded a panel of 14 candidate genes, to be examined in follow-up studies (Fig. 5.6.3 and Table 5.6.1).

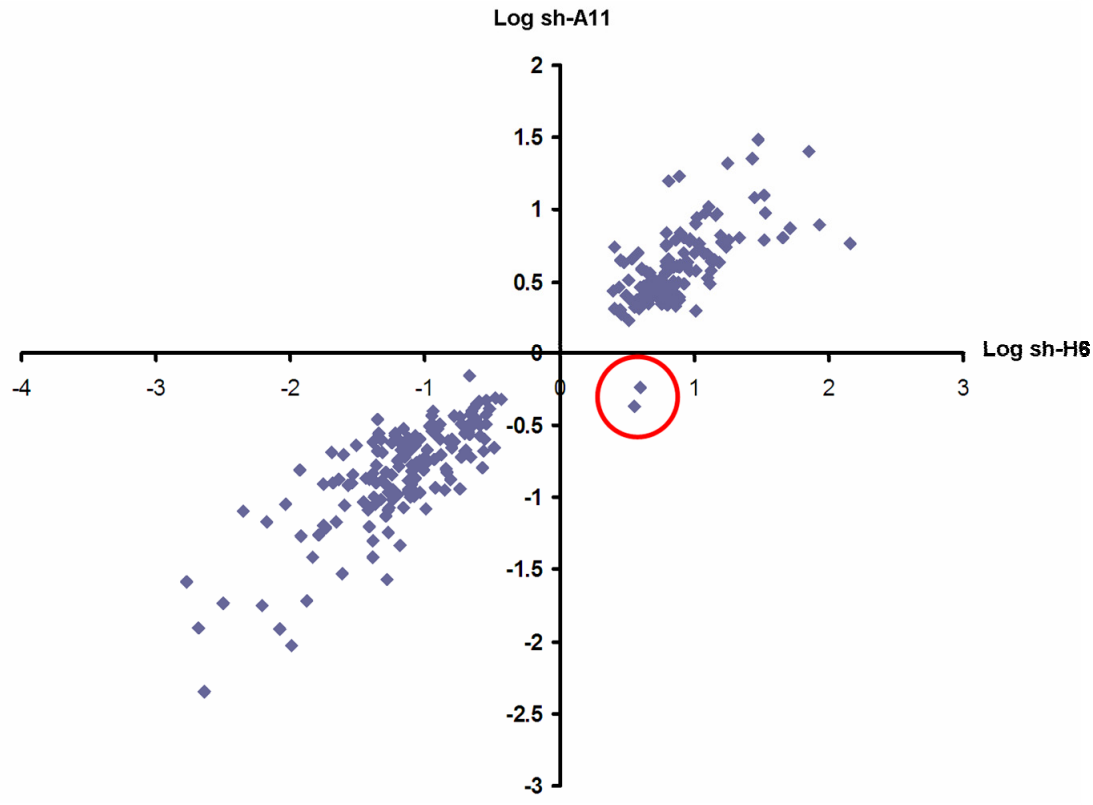


Figure 5.6.1: Graphic representation of the fold change log ratios for the 276 differentially expressed genes detected by transcription profiling of *CDKAL1* knock-down clones

The plot shows a high correlation between the fold change log ratios detected in the two sets of *CDKAL1* knock-down cell lines. The only two genes that show a different trend in the two knock-down clones are highlighted with a red circle.

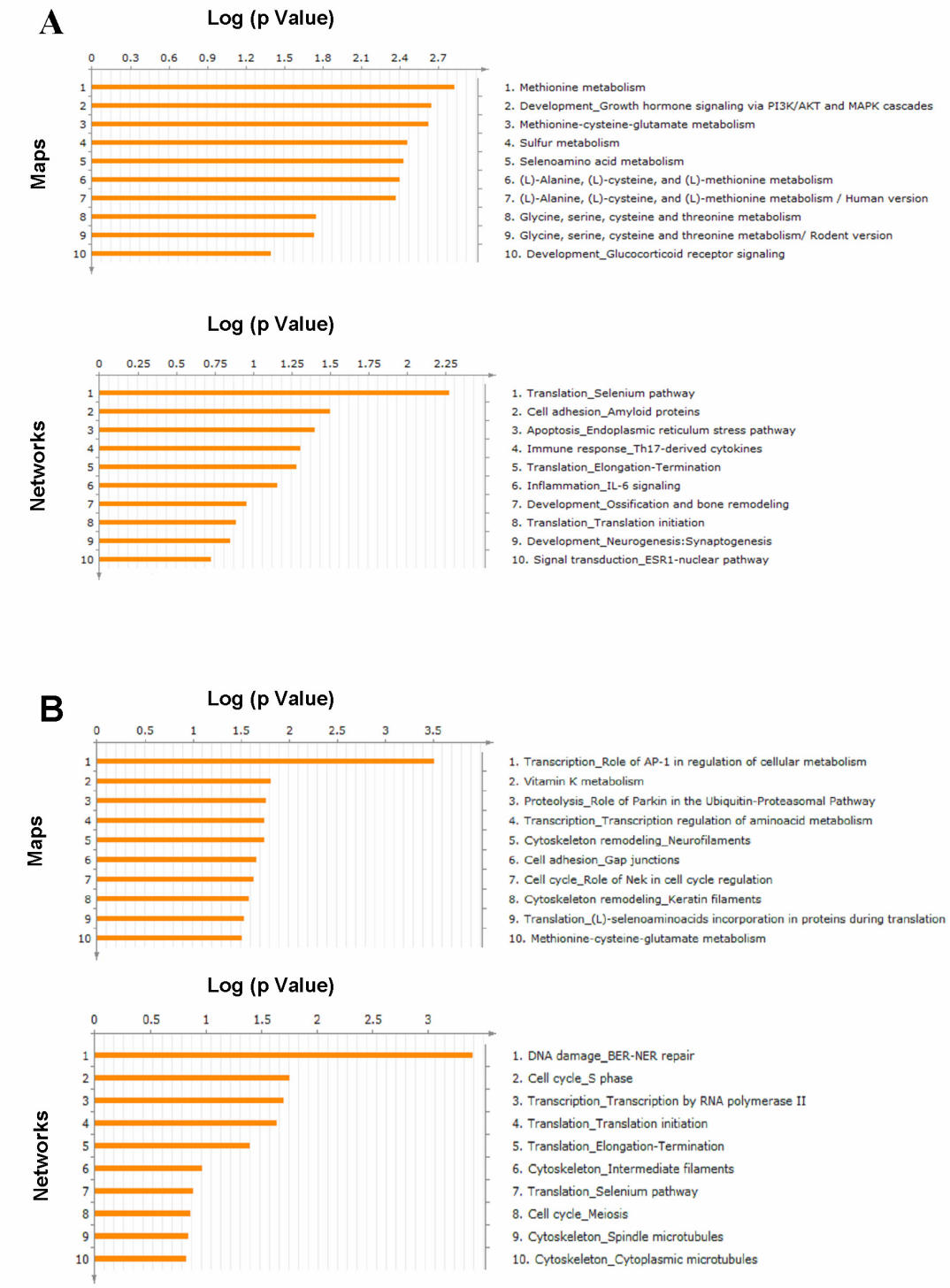

A
Log (p Value)
0 0.3 0.6 0.9 1.2 1.5 1.8 2.1 2.4 2.7
Maps
1. Methionine metabolism
2. Development_Growth hormone signaling via PI3K/AKT and MAPK cascades
3. Methionine-cysteine-glutamate metabolism
4. Sulfur metabolism
5. Selenoamino acid metabolism
6. (L)-Alanine, (L)-cysteine, and (L)-methionine metabolism
7. (L)-Alanine, (L)-cysteine, and (L)-methionine metabolism / Human version
8. Glycine, serine, cysteine and threonine metabolism
9. Glycine, serine, cysteine and threonine metabolism/ Rodent version
10. Development_Glucocorticoid receptor signaling
Log (p Value)
0 0.25 0.5 0.75 1 1.25 1.5 1.75 2 2.25
Networks
1. Translation_Selenium pathway
2. Cell adhesion_Amyloid proteins
3. Apoptosis_Endoplasmic reticulum stress pathway
4. Immune response_Th17-derived cytokines
5. Translation_Elongation-Termination
6. Inflammation_IL-6 signaling
7. Development_Ossification and bone remodeling
8. Translation_Translation initiation
9. Development_Neurogenesis:Synaptogenesis
10. Signal transduction_ESR1-nuclear pathway
B
Log (p Value)
0 0.5 1 1.5 2 2.5 3 3.5
Maps
1. Transcription_Role of AP-1 in regulation of cellular metabolism
2. Vitamin K metabolism
3. Proteolysis_Role of Parkin in the Ubiquitin-Proteasomal Pathway
4. Transcription_Transcription regulation of aminoacid metabolism
5. Cytoskeleton remodeling_Neurofilaments
6. Cell adhesion_Gap junctions
7. Cell cycle_Role of Nek in cell cycle regulation
8. Cytoskeleton remodeling_Keratin filaments
9. Translation_(L)-selenoaminoacids incorporation in proteins during translation
10. Methionine-cysteine-glutamate metabolism
Log (p Value)
0 0.5 1 1.5 2 2.5 3
Networks
1. DNA damage_BER-NER repair
2. Cell cycle_S phase
3. Transcription_Transcription by RNA polymerase II
4. Translation_Translation initiation
5. Translation_Elongation-Termination
6. Cytoskeleton_Intermediate filaments
7. Translation_Selenium pathway
8. Cell cycle_Meiosis
9. Cytoskeleton_Spindle microtubules
10. Cytoskeleton_Cytoplasmic microtubules

Figure 5.6.2: GeneGo pathway analysis of the differentially expressed genes detected by expression profiling of *CDKAL1* knock-down clones

Panel **A**: the top ten pathways identified through the analysis of the down-regulated gene list. Panel **B**: the top ten pathways identified through the analysis of the up-regulated gene list. Maps represent contained sets of experimentally validated interactions; networks refer to broader pathway reconstructions, which often include inferred interactions.

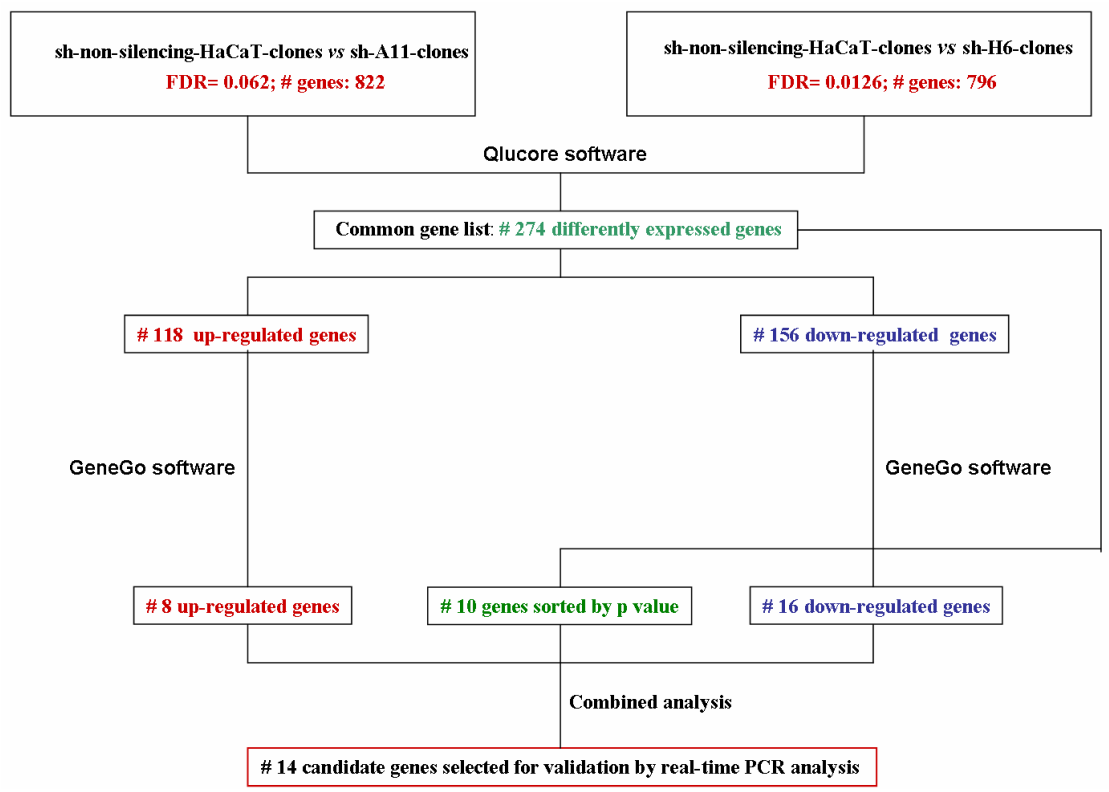


Figure 5.6.3: Schematic representation of the process leading to the selection of 14 genes to be followed-up

Table 5.6.1: Differentially expressed genes selected for follow-up

Up-regulated gene	Fold change log ratio	p value	Function of gene product *
GCLM	2.758	0.004	An enzyme that contributes to glutathione biosynthesis.
TUBB2C	2.663	0.0008	Microtubule constituent protein, which also binds GTP.
RPL29	2.633	5.22×10^{-7}	Ribosomal protein.
SEPX1	2.483	4.19×10^{-6}	Selenoprotein with methionine sulfoxide reductase activity.
POLR2L	2.44	1.68×10^{-7}	RNA polymerase II subunit.
RPL36A	2.408	1.09×10^{-5}	Ribosomal protein.
GTF2H5	2.043	1.14×10^{-8}	Subunit of the transcription/repair factor TFIIH.
NME1	1.772	9.52×10^{-8}	The RNA component of RNase MRP (RNase mitochondrial RNA processing), which is required for pre-rRNA processing and is essential for cell viability.
LSM10	1.714	2.90×10^{-9}	A protein consitutent of the U7 snRNP complex that is involved in histone 3'-end processing. It is required for cell cycle progression from G_1 to S phase.

*Function annotated in the Uniprot protein database.

Down-regulated gene	Fold change log ratio	p value	Function of gene product *
INHBE	-5.635	1.62×10^{-7}	A member of the activin beta family playing a role in pancreatic cell growth and proliferation.
TRIB3	-4.029	7.70×10^{-9}	A negative regulator of NF-κB and AKT1 signalling.
RIMS3	-3.482	9.19×10^{-8}	A protein involved in the regulation of synaptic membrane exocytosis.
DDIT3	-2.638	1.82×10^{-8}	A protein inhibiting the DNA-binding activity of the C/EBP and LAP transcription factors.
RAB3GAP1	-1.616	6.39×10^{-8}	Catalytic subunit of the RAB3 GTPase activating protein.

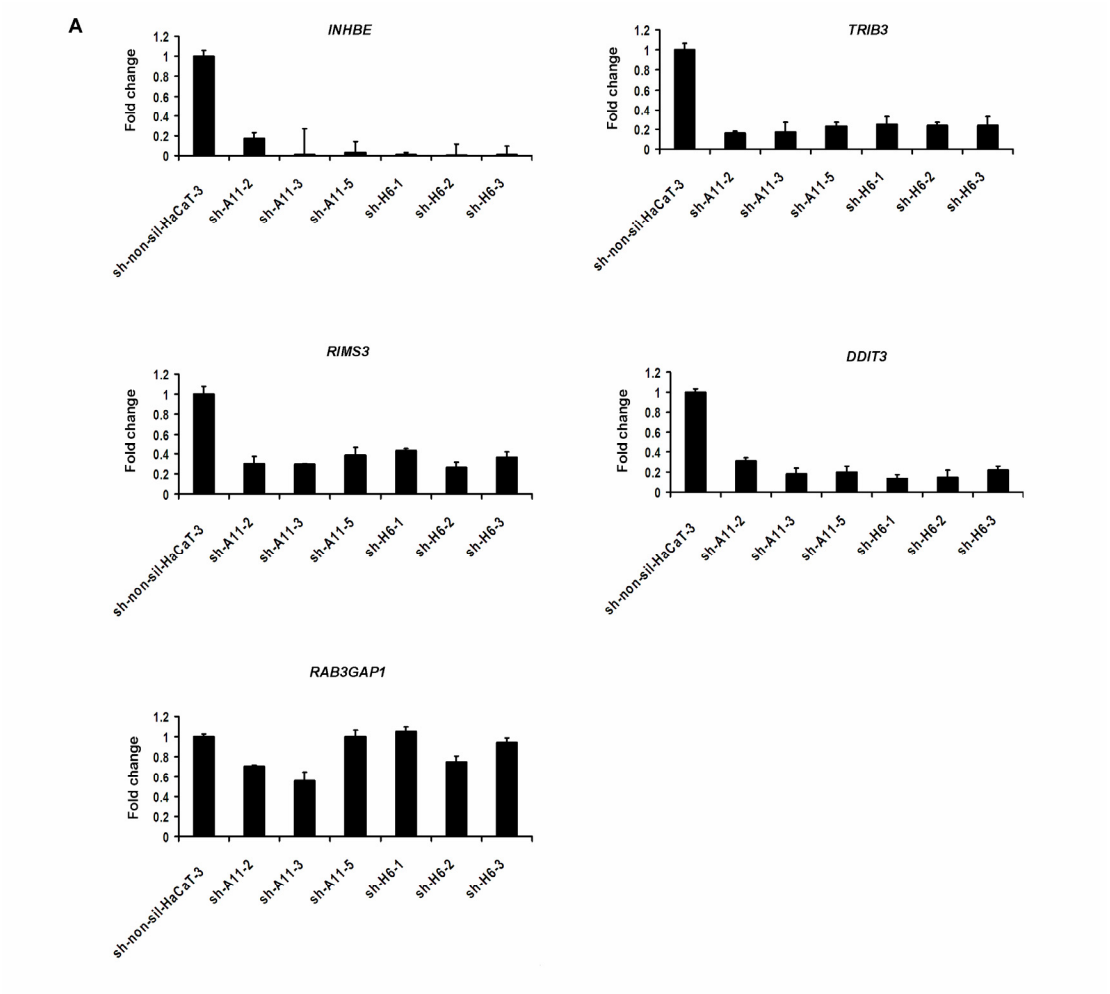
*Function annotated in the Uniprot protein database.

5.7 Follow-up of differentially expressed genes

In order to validate the differential expression of the 14 genes selected for follow-up, real-time PCRs were carried out in a panel of six knock-down clones, including the four that had been used in the microarray experiment (sh-A11-3; sh-A11-5; sh-H6-2 and sh-H6-3) and two additional lines (sh-A11-2 and sh-H6-1) identified during the screening of tGFP-expressing clones (Table 5.4.2). As shown in Figure 5.7.1, the trend in mRNA expression was confirmed for 13 out of the 14 genes examined.

These genes were further investigated by real-time PCR analysis of two knock-down Jurkat clones (sh-H6-7-C and sh-H6-8-D). As shown in Figure 5.7.2, the differential expression of the selected transcripts was not replicated in this cell type.

Finally, the disease relevance of the 13 genes was investigated by querying the psoriasis transcriptome dataset generated by the CASP consortium (Nair, Duffin et al. 2009). Gene expression data was available for 37 pairs of lesional and non-lesional patient skin biopsies and for 34 control samples, obtained from unaffected individuals. The analysis of this resource revealed two main gene clusters, defined according to the intensity of gene expression in the three sample classes (Fig. 5.7.3). Of note, evidence for over-expression in lesional skin was observed for 11 out of the 13 examined transcripts (all genes except *RPL36A* and *DDIT3*; Fig. 5.7.3).


A
INHBE
TRIB3
RIMS3
DDIT3
RAB3GAP1
Fold change
1.2
1
0.8
0.6
0.4
0.2
0
sh-non-sil-HaCaT-3
sh-A11-2
sh-A11-3
sh-A11-5
sh-H6-1
sh-H6-2
sh-H6-3

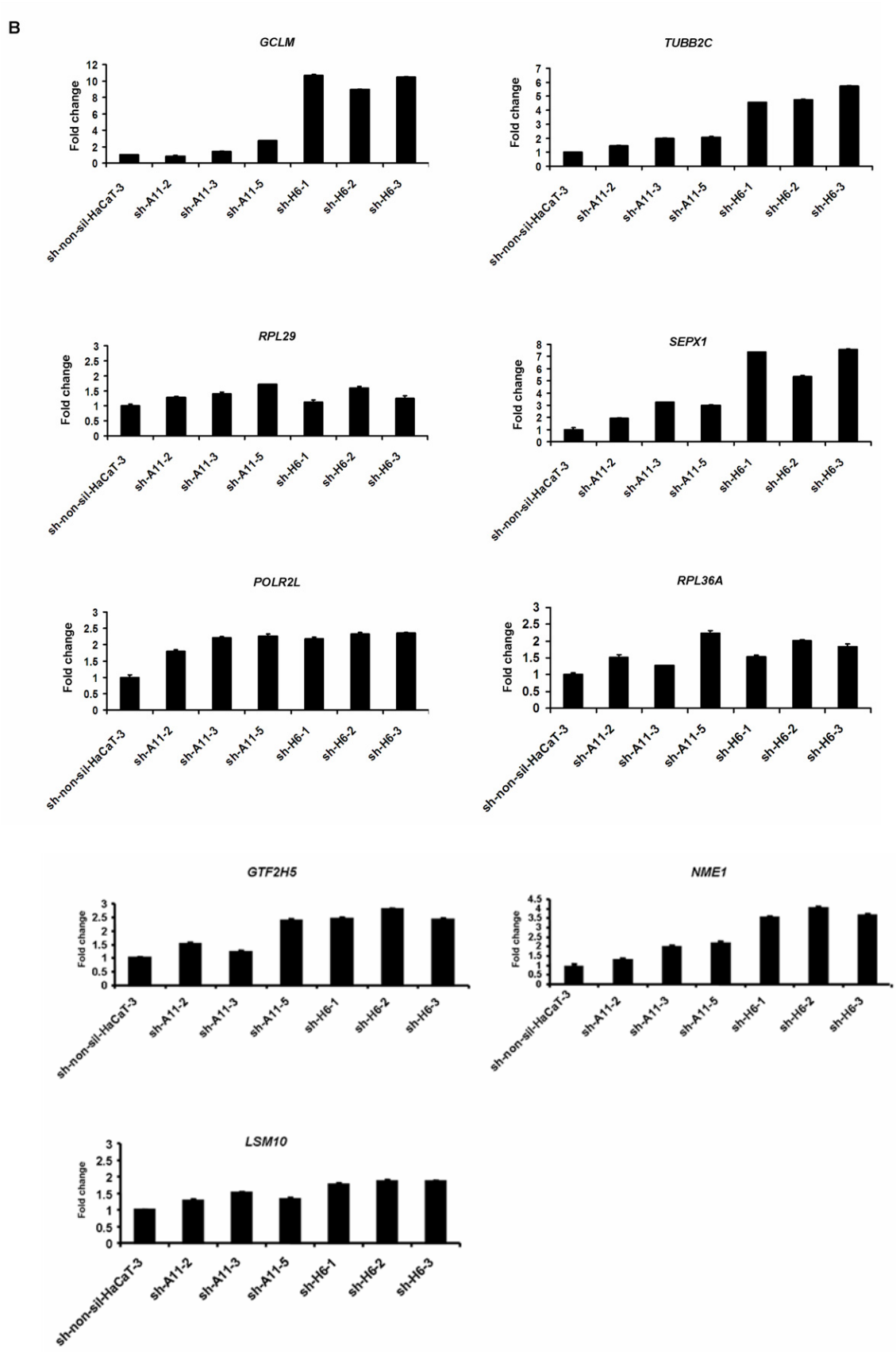
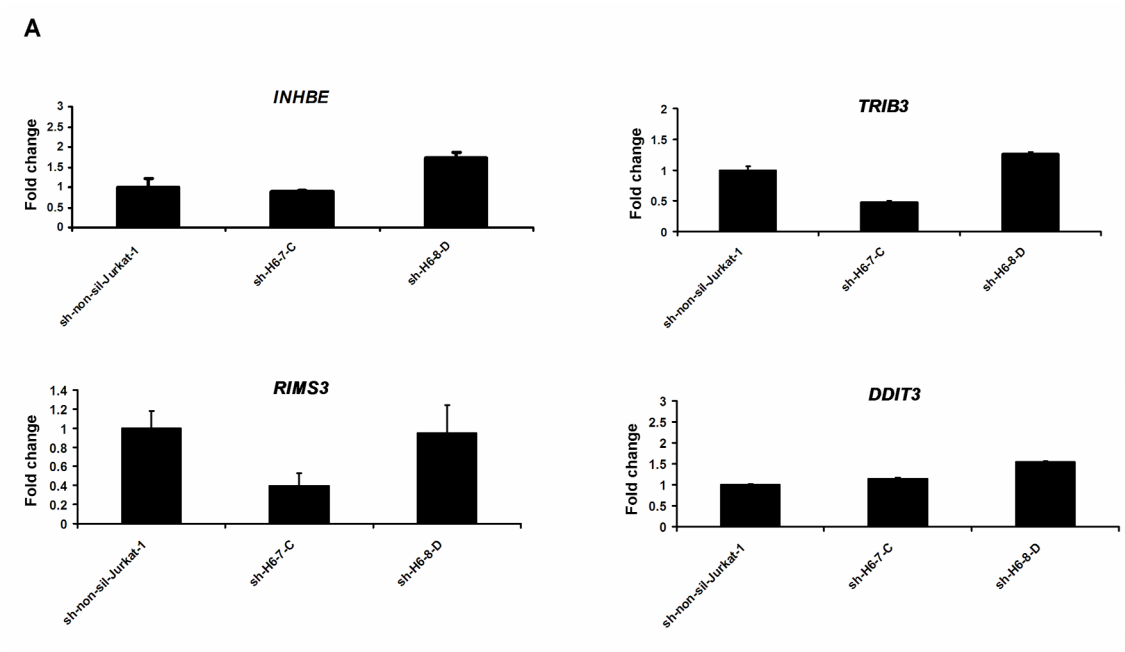

B
GCLM
TUBB2C
RPL29
SEPX1
POLR2L
RPL36A
GTF2H5
NME1
LSM10
Fold change
sh-non-sil-HaCaT-3
sh-A11-2
sh-A11-3
sh-A11-5
sh-H6-1
sh-H6-2
sh-H6-3

Figure 5.7.1: Follow-up of the differentially expressed genes identified by microarray analysis (HaCaT cells)

Transcript levels were quantified in six stably transduced HaCaT clones. All data were normalised to the ratio of *PPIA* and *RPLPO* expression. Panel **A**: real-time PCR analysis of the five down-regulated genes. Panel **B**: real-time PCR analysis of the nine up-regulated genes. Error bars refer to technical triplicates in real-time PCR experiments.


A
INHBE
Fold change
3
2.5
2
1.5
1
0.5
0
sh-non-sil-Jurkat-1
sh-H6-7-C
sh-H6-8-D
TRIB3
Fold change
2
1.5
1
0.5
0
sh-non-sil-Jurkat-1
sh-H6-7-C
sh-H6-8-D
RIMS3
Fold change
1.4
1.2
1
0.8
0.6
0.4
0.2
0
sh-non-sil-Jurkat-1
sh-H6-7-C
sh-H6-8-D
DDIT3
Fold change
3
2.5
2
1.5
1
0.5
0
sh-non-sil-Jurkat-1
sh-H6-7-C
sh-H6-8-D

B

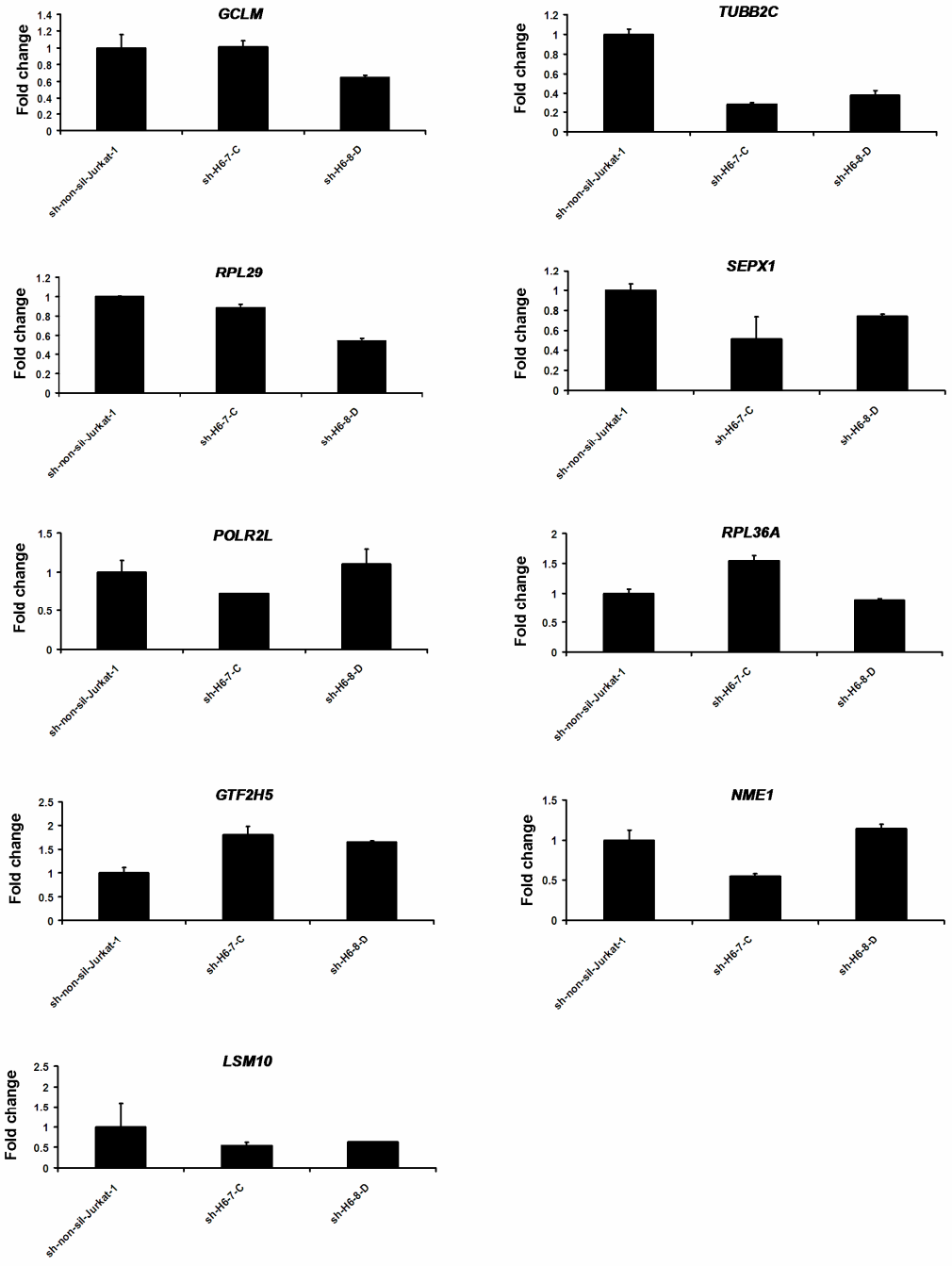

GCLM
Fold change
1.4
1.2
1
0.8
0.6
0.4
0.2
0
sh-non-sil-Jurkat-1
sh-H6-7-C
sh-H6-8-D
TUBB2C
Fold change
1.2
1
0.8
0.6
0.4
0.2
0
sh-non-sil-Jurkat-1
sh-H6-7-C
sh-H6-8-D
RPL29
Fold change
1.2
1
0.8
0.6
0.4
0.2
0
sh-non-sil-Jurkat-1
sh-H6-7-C
sh-H6-8-D
SEPX1
Fold change
1.2
1
0.8
0.6
0.4
0.2
0
sh-non-sil-Jurkat-1
sh-H6-7-C
sh-H6-8-D
POLR2L
Fold change
1.5
1
0.5
0
sh-non-sil-Jurkat-1
sh-H6-7-C
sh-H6-8-D
RPL36A
Fold change
2
1.5
1
0.5
0
sh-non-sil-Jurkat-1
sh-H6-7-C
sh-H6-8-D
GTF2H5
Fold change
2.5
2
1.5
1
0.5
0
sh-non-sil-Jurkat-1
sh-H6-7-C
sh-H6-8-D
NME1
Fold change
1.5
1
0.5
0
sh-non-sil-Jurkat-1
sh-H6-7-C
sh-H6-8-D
LSM10
Fold change
2.5
2
1.5
1
0.5
0
sh-non-sil-Jurkat-1
sh-H6-7-C
sh-H6-8-D

Figure 5.7.2: Follow-up of the differentially expressed genes identified by microarray analysis (Jurkat cells)

Transcript levels were quantified in two stably transduced Jurkat clones. All data were normalised to the ratio of *PPIA* and *RPLPO* expression. Panel **A**: real-time PCR analysis of the four validated down-regulated genes. Panel **B**: real-time PCR analysis of the nine validated up-regulated genes. Error bars refer to technical triplicates in real-time PCR experiments.

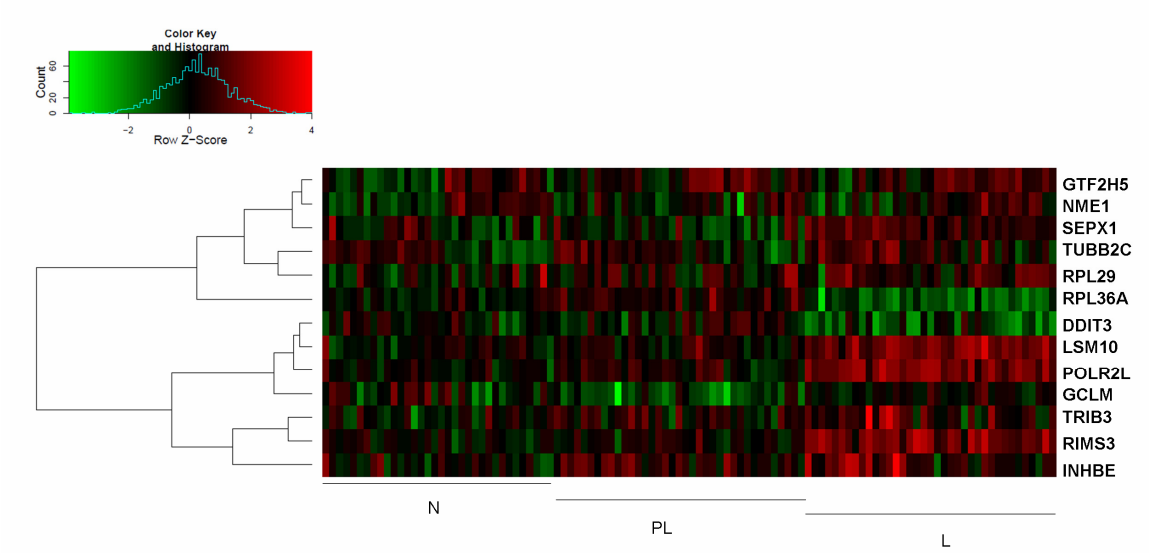


Figure 5.7.3: Hierarchical clustering analysis of CASP array data for the 13 differentially expressed genes

The genes were clustered in two groups, based on similarity in their expression profiles. The z-scores calculated for each gene are represented by colour intensities, as shown in the top left panel. N: normal skin samples obtained from unaffected individuals; PL: perilesional skin samples obtained from psoriatic patients; L: lesional skin samples obtained from psoriatic patients.

5.8 Discussion

The aim of these experiments was to explore the function of *CDKAL1*, with a view to understanding the mechanisms through which the gene contributes to disease susceptibility. The first phase of the study consisted of a series of bioinformatic searches. These indicated that CDKAL1 contains three highly conserved domains (a N-terminal MTTase motif, a central SAM and a C-terminal TRAM domain), which are typically found in proteins encoded by the methylthiotransferase (MTTase) gene family. These enzymes, which participate to the modification of tRNAs for ANN codons, can be further classified into five subfamilies (MiaB, RimO, MtaB, e-MtaB and MTL1) encoded by eukaryote and/or bacterial genomes. Bioinformatic analyses and functional assays have demonstrated that CDKAL1 belongs to the e-MtaB subfamily, which is found in higher eukaryotes and archaebacteria (Arragain, Handelman et al. 2010).

The next step in this work was the generation of the laboratory reagents required to investigate *CDKAL1* function at the experimental level. First, the specificity of a commercial α-CDKAL1 antibody was validated by assessing its ability to bind a recombinant GST-CDKAL1 protein and by carrying out western blot experiments in cells where the gene had been silenced by transient RNAi. Next, a panel of *CDKAL1* knock-down cells was generated by stable transduction of immortalized keratinocyte (HaCaT) and T lymphocyte (Jurkat) cell lines. These cell types were selected since they play a major role in the pathogenesis of psoriasis.

The characterization of the specificity of the α-CDKAL1 antibody and the generation of the stable *CDKAL1* knock-down cell lines allowed preliminary functional studies to be performed. The antibody was used in western blot and immunofluorescence experiments aimed at defining CDKAL1 subcellular localization. The results obtained with both approaches demonstrated that CDKAL1 is only found in the cell cytoplasm.

This is consistent with recently published data, showing that CDKAL1 localizes to the endoplasmic reticulum (ER) (Ohara-Imaizumi, Yoshida et al. 2010).

Previous studies have shown that *CDKAL1* lies in a chromosome region that is amplified in bladder tumours (Oeggerli, Schraml et al. 2006), suggesting the possibility that the gene may be involved in the control of cell cycle progression. To further explore this hypothesis, two independent assays were carried out in *CDKAL1* knock-down HaCaT cells. The WST-8 cell viability assay measures the activity of a mitochondrial dehydrogenase that is mostly functional when cells are proliferating. Conversely, PI staining determines the percentage of dividing cells within a given population, by measuring their DNA content. The results of both experiments support the notion that *CDKAL1* may contribute to the regulation of cell cycle progression and cell proliferation. In fact, the WST-8 assay showed that *CDKAL1* silencing results in a marked, time-dependent decrease in cell proliferation. At the same time, PI staining showed that *CDKAL1* knock-down induces growth arrest by increasing the proportion of cells in G_0/G_1 phase and decreasing the proportion of those in S phase.

To further investigate *CDKAL1* function, the expression profile of HaCaT knock-down cells was investigated by means of microarray analysis. To minimize the effects of clonal variability and sh-RNA sequence, the experiment was conducted on four independent clones, which had been transduced with two distinct sh-RNAs. The analysis of transcript levels yielded a panel of 14 candidate genes, potentially influenced by *CDKAL1* knock-down. Real-time PCR experiments, carried out in a total of six independent clones, subsequently validated the differential expression of 13 transcripts. An overview of these genes showed that the up-regulated transcripts mostly code for housekeeping proteins involved in the maintenance of cellular metabolism. Conversely, the down-regulated transcripts appear to participate in inflammatory processes. For

instance, *TRIB3* contributes to the regulation of NF-κB signalling by inhibiting the phosphorylation of its p65 subunit (Wu, Xu et al. 2003). Moreover *TRIB3*, *DDIT3* and *INHBE* all have a role in the unfolded protein response (Ohoka, Yoshii et al. 2005; Dombroski, Nayak et al. 2010), an ER-stress induced transcriptional program, which has been repeatedly linked to the onset of inflammation in T2D (Hotamisligil 2010) and CD (Kaser, Lee et al. 2008). Interestingly, a further down-regulated gene (*RIMS3*) has been implicated in molecular pathways associated with Ca^{2+} metabolism (Wang, Sugita et al. 2000). This is of some relevance, as it has been demonstrated that Ca^{2+} plays a critical role in regulating insulin release through the ATP cycle (Hughes, Lee et al. 2006). Moreover, a recent publication has shown that *CDKAL1* knock-down results in impaired insulin release and that this effect is mediated by an alteration of Ca^{2+} levels, due to reduced ATP generation (Ohara-Imaizumi, Yoshida et al. 2010).

The 13 genes that were differentially expressed in HaCaT cells were further investigated in *CDKAL1* knock-down Jurkat clones. Real-time PCR analysis of two independent cell lines failed to identify any evidence of differential expression for any of the examined genes. This discrepancy is likely to reflect the biological diversity of keratinocyte and T lymphocytes cell lines. In fact, a search of the BioGPS gene annotation portal (Wu, Orozco et al. 2009) revealed that T cells, unlike keratinocytes, express very low levels of the 13 critical genes. An additional aspect that needs to be considered is the lower knock-down efficiency of *CDKAL1* in Jurkat cell lines. Indeed, the 60% knock-down observed in these cells may not be sufficient to abolish *CDKAL1* function and affect the expression of downstream genes.

In the final phase of the study, the pathogenic relevance of the 13 differentially expressed genes was investigated by interrogating a publicly available transcriptome dataset. This led to the identification, of two clusters of transcripts sharing similar

expression profiles, an observation that further supports the existence of a functional connection between the differentially expressed genes. Eleven out of the 13 examined genes appeared to be over-expressed in lesional psoriatic skin, whereas two (*RPL36A* and *DDIT3*) were found to be down-regulated. These findings confirm the notion that the genes identified in this study are of genuine relevance to the pathogenesis of psoriasis.

Taken together, the results presented in this chapter indicate that *CDKAL1* is involved in a variety of cellular pathways, ranging from the maintenance of housekeeping functions (*e.g.* the post-transcriptional modification of tRNAs) to the orchestration of complex responses to environmental stimuli (*e.g.* the regulation of NF-κB signalling). In this context, *CDKAL1* is likely to contribute to disease onset, by disrupting cell cycle progression and inflammatory responses. Defining the mechanisms mediating these effects will require the implementation of further, hypothesis-driven functional studies.

Chapter 6: Discussion

The existence of shared genetic determinants contributing to the pathogenesis of multiple inflammatory disorders has been documented by the results of several GWASs (Seldin and Amos 2009). Thus, genetic studies are not only identifying novel disease pathways, but also uncovering unexpected connections between clinically distinct disorders (The Wellcome Trust Case-Control Consortium 2007). In this context, the aim of this project was to specifically investigate the genetic overlap between CD and psoriasis.

The notion that a number of shared genetic determinants may contribute to the pathogenesis of both conditions is supported by several lines of evidence. First of all, clinical observations suggest that individuals affected by CD may be at higher risk of developing psoriasis, compared to the general population (Najarian and Gottlieb 2003; Christophers 2007). Secondly, CD and psoriasis share some important immuno-pathogenic features, as both are caused by chronic inflammation of epithelial tissues, driven by IFN-γ producing Th1 cells and IL-17 producing Th17 cells (Di Cesare, Di Meglio et al. 2009; Strober and Fuss 2011). Finally, evidence for a genetic overlap between psoriasis and CD emerged from GWASs carried out before the onset of this study with SNPs in the *IL12B* and *IL23R* genes found to be significantly associated with both conditions (Duerr, Taylor et al. 2006; Capon, Di Meglio et al. 2007; Cargill, Schrodi et al. 2007; Barrett, Hansoul et al. 2008). In keeping with these results, a biologic drug targeting the protein encoded by *IL12B* has been found to be effective in the treatment of both CD and psoriasis (Toichi, Torres et al. 2006; Leonardi, Kimball et al. 2008; Papp, Langley et al. 2008; Ryan, Thrash et al. 2010).

To test the hypothesis that further CD markers may contribute to psoriasis susceptibility, a total of 26 CD loci were analysed in a sizeable case-control dataset. This led to the

identification of a novel disease association for a variant lying in intron 5 of the *CDKAL1* gene, a finding that was subsequently replicated in an independently ascertained dataset of North-American origin. Since these results were obtained, a number of psoriasis GWASs have been carried out (Zhang, Huang et al. 2009; Ellinghaus, Ellinghaus et al. 2010; Huffmeier, Uebe et al. 2010; Strange, Capon et al. 2010; Stuart, Nair et al. 2010; Sun, Cheng et al. 2010), but none has reported the identification of disease associated *CDKAL1* variants. This may reflect the experimental design of these studies, where only variants that exceed a given significance threshold in the discovery dataset are followed-up in replication samples. In fact, the GWAS carried out by Strange *et al.* (Strange, Capon et al. 2010) identified a disease association with a *CDKAL1* marker (rs7748720), but the statistical significance of this finding ($P = 6.3 \times 10^{-4}$) was not deemed sufficient to warrant a replication effort.

Of note, the GWAS carried out by Strange *et al.* (Strange, Capon et al. 2010) identified psoriasis associated alleles within the *REL* and *TYK2* genes (Strange, Capon et al. 2010), both of which are well known genetic determinants for CD (Franke, McGovern et al. 2010). Thus, a total of five CD genes have now been implicated in psoriasis susceptibility (*IL12B, IL23R, CDKAL1, REL* and *TYK2*), with more likely to emerge from an ongoing meta-analysis of two large studies (Strange, Capon et al. 2010; Stuart, Nair et al. 2010). Interestingly, *TYK2* encodes a protein kinase that mediates the signalling activity of the IL-23 receptor complex (Oyamada, Ikebe et al. 2009). The presence of associations at this locus further demonstrates that IL-23 driven activation of Th17 lymphocytes is the key pathogenic pathway shared by psoriasis and CD.

The *CDKAL1* SNP that has been associated with CD and psoriasis is completely independent from the markers conferring T2D susceptibility. Thus, *CDKAL1* must harbour two unrelated disease alleles. Although the pattern of LD conservation around

the critical SNPs indicates that both causal variants must lie within intron 5, their localization has not been refined any further. The use of the Immunochip array, which affords substantial coverage across the *CDKAL1* gene region, is expected to address this issue, as extended patients cohorts are currently being typed on this platform. Although these studies are likely to reduce the list of candidate susceptibility alleles to a small number of markers, the analysis non-coding SNPs associated with other common diseases suggests that the characterization of causal variants will be a challenging task, requiring the functional dissection of long-range *cis*-regulatory elements (Meyer, Maia et al. 2008; Wasserman, Aneas et al. 2010).

Despite the fact that the pathogenic involvement of *CDKAL1* is very well replicated (Saxena, Voight et al. 2007; Scott, Mohlke et al. 2007; Steinthorsdottir, Thorleifsson et al. 2007; The Wellcome Trust Case-Control Consortium 2007; Zeggini, Weedon et al. 2007; Barrett, Hansoul et al. 2008; Franke, McGovern et al. 2010; McGovern, Jones et al. 2010), the gene expression pattern has not been comprehensively investigated, with most studies focusing on the tissues that are affected by T2D (*e.g*. pancreas and skeletal muscle). Here, the analysis of *CDKAL1* expression has been extended to include a much wider range of cell types. This showed that *CDKAL1* transcripts are highly expressed in $CD4^+$ T cells but become down-regulated upon lymphocyte activation. The nature of the samples used in these experiments (pooled cDNAs obtained from cells stimulated with unspecific mitogen agents) precludes any particular speculation on the role of *CDKAL1* in lymphocyte activation. Further studies focusing on specific cell sub-populations (*e.g.* naïve *vs*. *in-vitro* differentiated T helper lymphocytes) and stimuli (*e.g.* direct activation of the T cell receptor) will enable researchers to address this issue, which is of considerable relevance, given the key role of $CD4^+$ T lymphocytes in the pathogenesis of psoriasis and CD.

The function of the CDKAL1 protein has also been the subject of little experimental investigation. Thus, the first step of the functional analyses reported in this thesis was the validation of appropriate laboratory reagents. The specificity of a commercial anti-CDKAL1 polyclonal antibody was validated, allowing the implementation of western blot and immunofluorescence protocols. These are fundamental applications, which, in the case of CDKAL1 have provided important information on the protein expression levels and subcellular localization. Further work will now be required to optimize the use of this antibody for other, equally important applications. For instance, the development of a protocol for immuno-histochemistry would be of particular interest, as it would allow researchers to investigate CDKAL1 expression in the tissues of patients affected by psoriasis, CD or T2D.

Another key reagent that has been generated in this study is a set of stable *CDKAL1* knock-down cells. The preliminary characterization of four indepenedent clones has generated some insights into the possible function of *CDKAL1*, as the results of transcriptome analyses suggested that the gene may be involved in the regulation of NF-κB mediated responses (through an effect on *TRIB3*) and Ca^{2+} metabolism (through an effect on *RIMS3*). The role of NF-κB signalling in the pathogenesis of chronic inflammation is well established (Barnes and Karin 1997; Hayden and Ghosh 2008) as is the link between Ca^{2+} signalling and cell proliferation (Roderick and Cook 2008). Thus, it is tempting to speculate that the dysregulation of the above pathways might be the mechanism whereby *CDKAL1* confers susceptibility to clinically distinct disorders. This pathogenic model can now be used to inform the design of further hypothesis-driven experiments (*e.g.* investigations of NF-κB signalling disruptions). Such studies could be carried out in the HaCaT cell lines described here, but also in knock-down clones derived from cell types that may have further relevance to the pathogenesis of

CD (*e.g.* the enterocyte-like Caco-2 human intestinal cell line) or T2D (*e.g*. the BxPC-3 human pancreatic cancer cell line). Cell lines stably over-expressing *CDKAL1* could also be generated and used to validate the findings obtained in knock-down systems.

Clearly, an in-depth investigation of *CDKAL1* function would require the establishment of an *in-vivo* knockout model. Of note a *Cdkal1* -/- mouse has been generated (Ohara-Imaizumi, Yoshida et al. 2010), but its phenotypic characterization has been restricted to the analysis of beta-pancreatic cells. Given the involvement of *CDKAL1* in other organ-specific pathologies, the analysis of other tissues and cell types would certainly be of considerable interest.

Taken together these results presented in this thesis show that the integration of large-scale association studies and *in-vitro* knockout technology has the potential to offer novel insights into pathogenetic processes. At the same time, the preliminary nature of the functional findings presented here underscores the arduous nature of the work that is necessary to translate the results of genetic studies into a detailed understanding of the molecular mechanisms underlying inherited disease susceptibility.

References

Abraham, C. and J. H. Cho (2009). "Inflammatory bowel disease." N Engl J Med **361**(21): 2066-2078.

Altshuler, D. M., R. A. Gibbs, et al. (2010). "Integrating common and rare genetic variation in diverse human populations." Nature **467**(7311): 52-58.

Anton, B. P., L. Saleh, et al. (2008). "RimO, a MiaB-like enzyme, methylthiolates the universally conserved Asp88 residue of ribosomal protein S12 in Escherichia coli." Proc Natl Acad Sci U S A **105**(6): 1826-1831.

Arragain, S., S. K. Handelman, et al. (2010). "Identification of eukaryotic and prokaryotic methylthiotransferase for biosynthesis of 2-methylthio-N6-threonylcarbamoyladenosine in tRNA." J Biol Chem **285**(37): 28425-28433.

Asumalahti, K., T. Laitinen, et al. (2000). "A candidate gene for psoriasis near HLA-C, HCR (Pg8), is highly polymorphic with a disease-associated susceptibility allele." Hum Mol Genet **9**(10): 1533-1542.

Asumalahti, K., C. Veal, et al. (2002). "Coding haplotype analysis supports HCR as the putative susceptibility gene for psoriasis at the MHC PSORS1 locus." Hum Mol Genet **11**(5): 589-597.

Auffray, C., M. H. Sieweke, et al. (2009). "Blood monocytes: development, heterogeneity, and relationship with dendritic cells." Annu Rev Immunol **27**: 669-692.

Australia and New Zealand Multiple Sclerosis Genetics Consortium (ANZgene) (2009). "Genome-wide association study identifies new multiple sclerosis susceptibility loci on chromosomes 12 and 20." Nat Genet **41**(7): 824-828.

Banchereau, J., F. Briere, et al. (2000). "Immunobiology of dendritic cells." Annu Rev Immunol **18**: 767-811.

Barker, J. N., R. S. Mitra, et al. (1991). "Keratinocytes as initiators of inflammation." Lancet **337**(8735): 211-214.

Barnes, P. J. and M. Karin (1997). "Nuclear factor-kappaB: a pivotal transcription factor in chronic inflammatory diseases." N Engl J Med **336**(15): 1066-1071.

Barrett, J. C., S. Hansoul, et al. (2008). "Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease." Nat Genet **40**(8): 955-962.

Beck, G. and G. S. Habicht (1996). "Immunity and the invertebrates." Sci Am **275**(5): 60-63, 66.

Benjamini, Y. (1995). "Controlling the false Discovery Rate: A pratical and Powerful approch to Multiple Testing." Journal of Royal Statistical Society **57**: 289-300.

Borgiani, P., L. Vallo, et al. (2002). "Exclusion of CARD15/NOD2 as a candidate susceptibility gene to psoriasis in the Italian population." Eur J Dermatol **12**(6): 540-542.

Bowcock, A. M. and J. G. Krueger (2005). "Getting under the skin: the immunogenetics of psoriasis." Nat Rev Immunol **5**(9): 699-711.

Brandrup, F., N. Holm, et al. (1982). "Psoriasis in monozygotic twins: variations in expression in individuals with identical genetic constitution." Acta Derm Venereol **62**(3): 229-236.

Burkitt, H. G. (2000). Wheater's functional histology: a text and colour atlas. Churchill Livingstone, Philadelphia (USA).

Burton, P. R., D. G. Clayton, et al. (2007). "Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants." Nat Genet **39**(11): 1329-1337.

Camp, R. D. R. (1998). Textbook of Dermatology. Blackwell Publishing LTD, Oxford (UK).

Capon, F., M. H. Allen, et al. (2004). "A synonymous SNP of the corneodesmosin gene leads to increased mRNA stability and demonstrates association with psoriasis across diverse ethnic groups." Hum Mol Genet **13**(20): 2361-2368.

Capon, F., M. J. Bijlmakers, et al. (2008). "Identification of ZNF313/RNF114 as a novel psoriasis susceptibility gene." Hum Mol Genet **17**(13): 1938-1945.

Capon, F., P. Di Meglio, et al. (2007). "Sequence variants in the genes for the interleukin-23 receptor (IL23R) and its ligand (IL12B) confer protection against psoriasis." Hum Genet **122**(2): 201-206.

Capon, F., M. Munro, et al. (2002). "Searching for the major histocompatibility complex psoriasis susceptibility gene." J Invest Dermatol **118**(5): 745-751.

Cargill, M., S. J. Schrodi, et al. (2007). "A large-scale genetic association study confirms IL12B and leads to the identification of IL23R as psoriasis-risk genes." Am J Hum Genet **80**(2): 273-290.

Cho, J. H. (2008). "The genetics and immunopathogenesis of inflammatory bowel disease." Nat Rev Immunol **8**(6): 458-466.

Christophers, E. (2001). "Psoriasis--epidemiology and clinical spectrum." Clin Exp Dermatol **26**(4): 314-320.

Christophers, E. (2007). "Comorbidities in psoriasis." Clin Dermatol **25**(6): 529-534.

Chu, C. C., P. Di Meglio, et al. (2011). "Harnessing dendritic cells in inflammatory skin diseases." Semin Immunol **23**(1): 28-41.

Clark, R. A. and T. S. Kupper (2006). "Misbehaving macrophages in the pathogenesis of psoriasis." J Clin Invest **116**(8): 2084-2087.

Cristopher, G. (2010). Rook's Textbook of Dermatology. Blackwell Publishing LTD, Oxford (UK).

de Cid, R., E. Riveira-Munoz, et al. (2009). "Deletion of the late cornified envelope

LCE3B and LCE3C genes as a susceptibility factor for psoriasis." Nat Genet **41**(2): 211-215.

Dendrou, C. A., V. Plagnol, et al. (2009). "Cell-specific protein phenotypes for the autoimmune locus IL2RA using a genotype-selectable human bioresource." Nat Genet **41**(9): 1011-1015.

Dharmani, P. and K. Chadee (2008). "Biologic therapies against inflammatory bowel disease: a dysregulated immune system and the cross talk with gastrointestinal mucosa hold the key." Curr Mol Pharmacol **1**(3): 195-212.

Di Cesare, A., P. Di Meglio, et al. (2009). "The IL-23/Th17 axis in the immunopathogenesis of psoriasis." J Invest Dermatol **129**(6): 1339-1350.

Di Meglio, P., A. Di Cesare, et al. (2011). "The IL23R R381Q gene variant protects against immune-mediated diseases by impairing IL-23-induced Th17 effector response in humans." PLoS One **6**(2): e17160.

Dombroski, B. A., R. R. Nayak, et al. (2010). "Gene expression and genetic variation in response to endoplasmic reticulum stress in human cells." Am J Hum Genet **86**(5): 719-729.

Dong, C. (2008). "TH17 cells in development: an updated view of their molecular identity and genetic programming." Nat Rev Immunol **8**(5): 337-348.

Duerr, R. H., K. D. Taylor, et al. (2006). "A genome-wide association study identifies IL23R as an inflammatory bowel disease gene." Science **314**(5804): 1461-1463.

Eisenbarth, S. C., D. A. Piggott, et al. (2003). "The master regulators of allergic inflammation: dendritic cells in Th2 sensitization." Curr Opin Immunol **15**(6): 620-626.

Elder, J. T., R. P. Nair, et al. (1994). "The genetics of psoriasis." Arch Dermatol **130**(2): 216-224.

Ellinghaus, E., D. Ellinghaus, et al. (2010). "Genome-wide association study identifies a psoriasis susceptibility locus at TRAF3IP2." Nat Genet **42**(11): 991-995.

Enerback, C., F. Enlund, et al. (2000). "S gene (Corneodesmosin) diversity and its relationship to psoriasis; high content of cSNP in the HLA-linked S gene." J Invest Dermatol **114**(6): 1158-1163.

Enerback, C., S. Nilsson, et al. (2000). "Stronger association with HLA-Cw6 than with corneodesmosin (S-gene) polymorphisms in Swedish psoriasis patients." Arch Dermatol Res **292**(11): 525-530.

Falk, K., O. Rotzschke, et al. (1993). "Allele-specific peptide ligand motifs of HLA-C molecules." Proc Natl Acad Sci U S A **90**(24): 12005-12009.

Fan, X., S. Yang, et al. (2008). "Fine mapping of the psoriasis susceptibility locus PSORS1 supports HLA-C as the susceptibility gene in the Han Chinese population." PLoS Genet **4**(3): e1000038.

Fisher, S. A., M. Tremelling, et al. (2008). "Genetic determinants of ulcerative colitis include the ECM1 locus and five loci implicated in Crohn's disease." Nat Genet **40**(6): 710-712.

Franke, A., J. Hampe, et al. (2007). "Systematic association mapping identifies NELL1 as a novel IBD disease gene." PLoS One **2**(1): e691.

Franke, A., D. P. McGovern, et al. (2010). "Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci." Nat Genet **42**(12): 1118-1125.

Frazer, K. A., D. G. Ballinger, et al. (2007). "A second generation human haplotype map of over 3.1 million SNPs." Nature **449**(7164): 851-861.

Fukata, M., A. S. Vamadevan, et al. (2009). "Toll-like receptors (TLRs) and Nod-like receptors (NLRs) in inflammatory disorders." Semin Immunol **21**(4): 242-253.

Ganguly, D., G. Chamilos, et al. (2009). "Self-RNA-antimicrobial peptide complexes activate human dendritic cells through TLR7 and TLR8." J Exp Med **206**(9): 1983-1994.

Gateva, V., J. K. Sandling, et al. (2009). "A large-scale replication study identifies TNIP1, PRDM1, JAZF1, UHRF1BP1 and IL10 as risk loci for systemic lupus erythematosus." Nat Genet **41**(11): 1228-1233.

Gelfand, J. M., A. L. Neimann, et al. (2006). "Risk of myocardial infarction in patients with psoriasis." JAMA **296**(14): 1735-1741.

Gottlieb, A., N. J. Korman, et al. (2008). "Guidelines of care for the management of psoriasis and psoriatic arthritis: Section 2. Psoriatic arthritis: overview and guidelines of care for treatment with an emphasis on the biologics." J Am Acad Dermatol **58**(5): 851-864.

Gottlieb, A. B. and F. Dann (2009). "Comorbidities in patients with psoriasis." Am J Med **122**(12): 1150 e1151-1159.

Graham, R. R., C. Cotsapas, et al. (2008). "Genetic variants near TNFAIP3 on 6q23 are associated with systemic lupus erythematosus." Nat Genet **40**(9): 1059-1061.

Gregersen, P. K., C. I. Amos, et al. (2009). "REL, encoding a member of the NF-kappaB family of transcription factors, is a newly defined risk locus for rheumatoid arthritis." Nat Genet **41**(7): 820-823.

Griffiths, C. E. and J. N. Barker (2007). "Pathogenesis and clinical features of psoriasis." Lancet **370**(9583): 263-271.

Griffiths, C. E., E. Christophers, et al. (2007). "A classification of psoriasis vulgaris according to phenotype." Br J Dermatol **156**(2): 258-262.

Hampe, J., A. Franke, et al. (2007). "A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in

ATG16L1." Nat Genet **39**(2): 207-211.

Han, J. W., H. F. Zheng, et al. (2009). "Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus." Nat Genet **41**(11): 1234-1237.

Hanauer, S. B. (2009). Inflammatory bowel diseases On-Line Textbook ACP Medicine.

Hansen, T. H. and M. Bouvier (2009). "MHC class I antigen presentation: learning from viral evasion strategies." Nat Rev Immunol **9**(7): 503-513.

Harder, J. and G. Nunez (2009). "Functional expression of the intracellular pattern recognition receptor NOD1 in human keratinocytes." J Invest Dermatol **129**(5): 1299-1302.

Hayden, M. S. and S. Ghosh (2008). "Shared principles in NF-kappaB signaling." Cell **132**(3): 344-362.

Hollox, E. J., U. Huffmeier, et al. (2008). "Psoriasis is associated with increased beta-defensin genomic copy number." Nat Genet **40**(1): 23-25.

Hotamisligil, G. S. (2010). "Endoplasmic reticulum stress and the inflammatory basis of metabolic disease." Cell **140**(6): 900-917.

Howie, B. N., P. Donnelly, et al. (2009). "A flexible and accurate genotype imputation method for the next generation of genome-wide association studies." PLoS Genet **5**(6): e1000529.

Huffmeier, U., S. Uebe, et al. (2010). "Common variants at TRAF3IP2 are associated with susceptibility to psoriatic arthritis and psoriasis." Nat Genet **42**(11): 996-999.

Hughes, E., A. K. Lee, et al. (2006). "Dominant role of sarcoendoplasmic reticulum Ca2+-ATPase pump in Ca2+ homeostasis and exocytosis in rat pancreatic beta-cells." Endocrinology **147**(3): 1396-1407.

Hugot, J. P., M. Chamaillard, et al. (2001). "Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease." Nature **411**(6837): 599-603.

Hundhausen, C., A. Bertoni, et al. (2011). "Allele-Specific Cytokine Responses at the HLA-C Locus: Implications for Psoriasis." J Invest Dermatol. Epub 24 Nov

International Psoriasis Genetics Consortium (2003). "The International Psoriasis Genetics Study: assessing linkage to 14 candidate susceptibility loci in a cohort of 942 affected sib pairs." Am J Hum Genet **73**(2): 430-437.

Irizarry, R. A., B. Hobbs, et al. (2003). "Exploration, normalization, and summaries of high density oligonucleotide array probe level data." Biostatistics **4**(2): 249-264.

Kagami, S., H. L. Rizzo, et al. (2010). "Circulating Th17, Th22, and Th1 cells are increased in psoriasis." J Invest Dermatol **130**(5): 1373-1383.

Kaplan, D. H., M. C. Jenison, et al. (2005). "Epidermal langerhans cell-deficient mice develop enhanced contact hypersensitivity." Immunity **23**(6): 611-620.

Kaser, A., A. H. Lee, et al. (2008). "XBP1 links ER stress to intestinal inflammation and confers genetic risk for human inflammatory bowel disease." Cell **134**(5): 743-756.

Kawai, T. and S. Akira (2008). "Toll-like receptor and RIG-I-like receptor signaling." Ann N Y Acad Sci **1143**: 1-20.

Lande, R., J. Gregorio, et al. (2007). "Plasmacytoid dendritic cells sense self-DNA coupled with antimicrobial peptide." Nature **449**(7162): 564-569.

Lander, E. S., L. M. Linton, et al. (2001). "Initial sequencing and analysis of the human genome." Nature **409**(6822): 860-921.

Lawrence, M. B. and T. A. Springer (1991). "Leukocytes roll on a selectin at physiologic flow rates: distinction from and prerequisite for adhesion through integrins." Cell **65**(5): 859-873.

Leonardi, C. L., A. B. Kimball, et al. (2008). "Efficacy and safety of ustekinumab, a human interleukin-12/23 monoclonal antibody, in patients with psoriasis: 76-week results from a randomised, double-blind, placebo-controlled trial (PHOENIX 1)." Lancet **371**(9625): 1665-1674.

Levine, B. and V. Deretic (2007). "Unveiling the roles of autophagy in innate and adaptive immunity." Nat Rev Immunol **7**(10): 767-777.

Li, Y., C. J. Willer, et al. (2010). "MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes." Genet Epidemiol **34**(8): 816-834.

Liang, S. C., X. Y. Tan, et al. (2006). "Interleukin (IL)-22 and IL-17 are coexpressed by Th17 cells and cooperatively enhance expression of antimicrobial peptides." J Exp Med **203**(10): 2271-2279.

Libioulle, C., E. Louis, et al. (2007). "Novel Crohn disease locus identified by genome-wide association maps to a gene desert on 5p13.1 and modulates expression of PTGER4." PLoS Genet **3**(4): e58.

Livak, K. J. and T. D. Schmittgen (2001). "Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method." Methods **25**(4): 402-408.

Lopez-Albaitero, A., R. Mailliard, et al. (2009). "Maturation pathways of dendritic cells determine TAP1 and TAP2 levels and cross-presenting function." J Immunother **32**(5): 465-473.

Maher, B. (2008). "Personal genomes: The case of the missing heritability." Nature **456**(7218): 18-21.

Mallon, E., M. Bunce, et al. (1997). "HLA-CW*0602 is a susceptibility factor in type I psoriasis, and evidence Ala-73 is increased in male type I psoriatics." J Invest Dermatol **109**(2): 183-186.

Manabu, F. (2010). "Regulatory B cells in skin and connective tissue disease." Journal of Dermatological Science **60**: 1-7.

Mantovani, A., A. Sica, et al. (2004). "The chemokine system in diverse forms of macrophage activation and polarization." Trends Immunol **25**(12): 677-686.

Martinez, G. J., R. I. Nurieva, et al. (2008). "Regulation and function of proinflammatory TH17 cells." Ann N Y Acad Sci **1143**: 188-211.

Mathew, C. G. and C. M. Lewis (2004). "Genetics of inflammatory bowel disease: progress and prospects." Hum Mol Genet **13 Spec No 1**: R161-168.

Mayer, G. (2006). Microbiology and Immunology On-Line Textbook, USC School of Medicine,UCSF.

McDonald, C., N. Inohara, et al. (2005). "Peptidoglycan signaling in innate immunity and inflammatory disease." J Biol Chem **280**(21): 20177-20180.

McGovern, D. P., M. R. Jones, et al. (2010). "Fucosyltransferase 2 (FUT2) non-secretor status is associated with Crohn's disease." Hum Mol Genet **19**(17): 3468-3476.

Menter, A., A. Gottlieb, et al. (2008). "Guidelines of care for the management of psoriasis and psoriatic arthritis: Section 1. Overview of psoriasis and guidelines of care for the treatment of psoriasis with biologics." J Am Acad Dermatol **58**(5): 826-850.

Meyer, K. B., A. T. Maia, et al. (2008). "Allele-specific up-regulation of FGFR2 increases susceptibility to breast cancer." PLoS Biol **6**(5): e108.

Miller, L. S. and R. L. Modlin (2007). "Human keratinocyte Toll-like receptors promote distinct immune responses." J Invest Dermatol **127**(2): 262-263.

Miller, L. S. and R. L. Modlin (2007). "Toll-like receptors in the skin." Semin Immunopathol **29**(1): 15-26.

Miller, S. A., D. D. Dykes, et al. (1988). "A simple salting out procedure for extracting

DNA from human nucleated cells." Nucleic Acids Res **16**(3): 1215.

Moser, M. and K. M. Murphy (2000). "Dendritic cell regulation of TH1-TH2 development." Nat Immunol **1**(3): 199-205.

Muhlrad, D. and R. Parker (1999). "Aberrant mRNAs with extended 3' UTRs are substrates for rapid degradation by mRNA surveillance." RNA **5**(10): 1299-1307.

Musone, S. L., K. E. Taylor, et al. (2008). "Multiple polymorphisms in the TNFAIP3 region are independently associated with systemic lupus erythematosus." Nat Genet **40**(9): 1062-1064.

Nair, R. P., K. C. Duffin, et al. (2009). "Genome-wide scan reveals association of psoriasis with IL-23 and NF-kappaB pathways." Nat Genet **41**(2): 199-204.

Nair, R. P., T. Henseler, et al. (1997). "Evidence for two psoriasis susceptibility loci (HLA and 17q) and two novel candidate regions (16q and 20p) by genome-wide scan." Hum Mol Genet **6**(8): 1349-1356.

Nair, R. P., P. Stuart, et al. (2001). "Lack of association between NOD2 3020InsC frameshift mutation and psoriasis." J Invest Dermatol **117**(6): 1671-1672.

Nair, R. P., P. E. Stuart, et al. (2006). "Sequence and haplotype analysis supports HLA-C as the psoriasis susceptibility 1 gene." Am J Hum Genet **78**(5): 827-851.

Najarian, D. J. and A. B. Gottlieb (2003). "Connections between psoriasis and Crohn's disease." J Am Acad Dermatol **48**(6): 805-821; quiz 822-804.

Nakhaei, P., P. Genin, et al. (2009). "RIG-I-like receptors: sensing and responding to RNA virus infection." Semin Immunol **21**(4): 215-222.

Nejentsev, S., N. Walker, et al. (2009). "Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes." Science **324**(5925): 387-389.

Nestle, F. O., C. Conrad, et al. (2005). "Plasmacytoid predendritic cells initiate psoriasis through interferon-alpha production." J Exp Med **202**(1): 135-143.

Nestle, F. O., P. Di Meglio, et al. (2009). "Skin immune sentinels in health and disease." Nat Rev Immunol **9**(10): 679-691.

Nestle, F. O. and M. Gilliet (2005). "Defining upstream elements of psoriasis pathogenesis: an emerging role for interferon alpha." J Invest Dermatol **125**(5): xiv-xv.

Nestle, F. O., D. H. Kaplan, et al. (2009). "Psoriasis." N Engl J Med **361**(5): 496-509.

Nestle, F. O., L. A. Turka, et al. (1994). "Characterization of dermal dendritic cells in psoriasis. Autostimulation of T lymphocytes and induction of Th1 type cytokines." J Clin Invest **94**(1): 202-209.

Nickoloff, B. J. (1999). "Skin innate immune system in psoriasis: friend or foe?" J Clin Invest **104**(9): 1161-1164.

Oeggerli, M., P. Schraml, et al. (2006). "E2F3 is the main target gene of the 6p22 amplicon with high specificity for human bladder cancer." Oncogene **25**(49): 6538-6543.

Ogura, Y., D. K. Bonen, et al. (2001). "A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease." Nature **411**(6837): 603-606.

Ohara-Imaizumi, M., M. Yoshida, et al. (2010). "Deletion of CDKAL1 affects mitochondrial ATP generation and first-phase insulin exocytosis." PLoS One **5**(12): e15553.

Ohoka, N., S. Yoshii, et al. (2005). "TRB3, a novel ER stress-inducible gene, is induced via ATF4-CHOP pathway and is involved in cell death." EMBO J **24**(6): 1243-1255.

Oyamada, A., H. Ikebe, et al. (2009). "Tyrosine kinase 2 plays critical roles in the

pathogenic CD4 T cell responses for the development of experimental autoimmune encephalomyelitis." J Immunol **183**(11): 7539-7546.

Papp, K. A., R. G. Langley, et al. (2008). "Efficacy and safety of ustekinumab, a human interleukin-12/23 monoclonal antibody, in patients with psoriasis: 52-week results from a randomised, double-blind, placebo-controlled trial (PHOENIX 2)." Lancet **371**(9625): 1675-1684.

Parkes, M., J. C. Barrett, et al. (2007). "Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility." Nat Genet **39**(7): 830-832.

Pearson, T. A. and T. A. Manolio (2008). "How to interpret a genome-wide association study." JAMA **299**(11): 1335-1344.

Penn, L., M. White, et al. (2009). "Prevention of type 2 diabetes in adults with impaired glucose tolerance: the European Diabetes Prevention RCT in Newcastle upon Tyne, UK." BMC Public Health **9**: 342.

Pickering, B. M. and A. E. Willis (2005). "The implications of structured 5' untranslated regions on translation and disease." Semin Cell Dev Biol **16**(1): 39-47.

Plenge, R. M., C. Cotsapas, et al. (2007). "Two independent alleles at 6q23 associated with risk of rheumatoid arthritis." Nat Genet **39**(12): 1477-1482.

Podolsky, D. K. (2002). "Inflammatory bowel disease." N Engl J Med **347**(6): 417-429.

Power, C. and J. Elliott (2006). "Cohort profile: 1958 British birth cohort (National Child Development Study)." Int J Epidemiol **35**(1): 34-41.

Purcell, S., B. Neale, et al. (2007). "PLINK: a tool set for whole-genome association and population-based linkage analyses." Am J Hum Genet **81**(3): 559-575.

Raelson, J. V., R. D. Little, et al. (2007). "Genome-wide association study for Crohn's disease in the Quebec Founder Population identifies multiple validated disease

loci." Proc Natl Acad Sci U S A **104**(37): 14747-14752.

Rioux, J. D., M. J. Daly, et al. (2001). "Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn disease." Nat Genet **29**(2): 223-228.

Rioux, J. D., R. J. Xavier, et al. (2007). "Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis." Nat Genet **39**(5): 596-604.

Risch, N. J. (2000). "Searching for genetic determinants in the new millennium." Nature **405**(6788): 847-856.

Roberson, E. D. and A. M. Bowcock (2010). "Psoriasis genetics: breaking the barrier." Trends Genet **26**(9): 415-423.

Roderick, H. L. and S. J. Cook (2008). "Ca2+ signalling checkpoints in cancer: remodelling Ca2+ for cancer cell proliferation and survival." Nat Rev Cancer **8**(5): 361-375.

Ryan, C., B. Thrash, et al. (2010). "The use of ustekinumab in autoimmune disease." Expert Opin Biol Ther **10**(4): 587-604.

Satsangi, J., M. S. Silverberg, et al. (2006). "The Montreal classification of inflammatory bowel disease: controversies, consensus, and implications." Gut **55**(6): 749-753.

Saxena, R., B. F. Voight, et al. (2007). "Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels." Science **316**(5829): 1331-1336.

Scott, L. J., K. L. Mohlke, et al. (2007). "A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants." Science **316**(5829): 1341-1345.

Seldin, M. F. and C. I. Amos (2009). "Shared susceptibility variations in autoimmune diseases: a brief perspective on common issues." Genes Immun **10**(1): 1-4.

Serbina, N. V., T. P. Salazar-Mather, et al. (2003). "TNF/iNOS-producing dendritic cells mediate innate immune defense against bacterial infection." Immunity **19**(1): 59-70.

Shapiro, J., A. D. Cohen, et al. (2007). "The association between psoriasis, diabetes mellitus, and atherosclerosis in Israel: a case-control study." J Am Acad Dermatol **56**(4): 629-634.

Sigurdsson, S., G. Nordmark, et al. (2005). "Polymorphisms in the tyrosine kinase 2 and interferon regulatory factor 5 genes are associated with systemic lupus erythematosus." Am J Hum Genet **76**(3): 528-537.

Simon, M., N. Jonca, et al. (2001). "Refined characterization of corneodesmosin proteolysis during terminal differentiation of human epidermis and its relationship to desquamation." J Biol Chem **276**(23): 20292-20299.

Smith, P. K., R. I. Krohn, et al. (1985). "Measurement of protein using bicinchoninic acid." Anal Biochem **150**(1): 76-85.

Smyth, D. J., J. D. Cooper, et al. (2006). "A genome-wide association study of nonsynonymous SNPs identifies a type 1 diabetes locus in the interferon-induced helicase (IFIH1) region." Nat Genet **38**(6): 617-619.

Smyth, G. K. (2004). "Linear models and empirical bayes methods for assessing differential expression in microarray experiments." Stat Appl Genet Mol Biol **3**: Article3.

Steinthorsdottir, V., G. Thorleifsson, et al. (2007). "A variant in CDKAL1 influences insulin response and risk of type 2 diabetes." Nat Genet **39**(6): 770-775.

Strachan, T., Read Andrew. P. (2004). Human Molecular Genetics. Garland Publishing, New York (USA).

Strange, A., F. Capon, et al. (2010). "A genome-wide association study identifies new

psoriasis susceptibility loci and an interaction between HLA-C and ERAP1." Nat Genet **42**(11): 985-990.

Stratis, A., M. Pasparakis, et al. (2006). "Pathogenic role for skin macrophages in a mouse model of keratinocyte-induced psoriasis-like skin inflammation." J Clin Invest **116**(8): 2094-2104.

Strober, W. and I. J. Fuss (2011). "Proinflammatory cytokines in the pathogenesis of inflammatory bowel diseases." Gastroenterology **140**(6): 1756-1767 e1751.

Strober, W., I. J. Fuss, et al. (2002). "The immunology of mucosal models of inflammation." Annu Rev Immunol **20**: 495-549.

Stuart, P. E., R. P. Nair, et al. (2010). "Genome-wide association analysis identifies three psoriasis susceptibility loci." Nat Genet **42**(11): 1000-1004.

Sun, L. D., H. Cheng, et al. (2010). "Association analyses identify six new psoriasis susceptibility loci in the Chinese population." Nat Genet **42**(11): 1005-1009.

Tazi Ahnini, R., N. J. Camp, et al. (1999). "Novel genetic association between the corneodesmosin (MHC S) gene and susceptibility to psoriasis." Hum Mol Genet **8**(6): 1135-1140.

Teng, R., K. Johkura, et al. (2003). "Morphological analysis of leucocyte transmigration in the pleural cavity." J Anat **203**(4): 391-404.

The 1000 Genome Project Consortium (2010). "A map of human genome variation from population-scale sequencing." Nature **467**(7319): 1061-1073.

The International HapMap Consortium (2005). "A haplotype map of the human genome." Nature **437**(7063): 1299-1320.

The Wellcome Trust Case-Control Consortium (2007). "Genome-wide association study of 14000 cases of seven common diseases and 3000 shared controls." Nature **447**: 661-678.

Thomson, W., A. Barton, et al. (2007). "Rheumatoid arthritis association at 6q23." Nat Genet **39**(12): 1431-1433.

Thorisson, G. A., A. V. Smith, et al. (2005). "The International HapMap Project Web site." Genome Res **15**(11): 1592-1593.

Toichi, E., G. Torres, et al. (2006). "An anti-IL-12p40 antibody down-regulates type 1 cytokines, chemokines, and IL-12/IL-23 in psoriasis." J Immunol **177**(7): 4917-4926.

Trembath, R. C., R. L. Clough, et al. (1997). "Identification of a major susceptibility locus on chromosome 6p and evidence for further disease loci revealed by a two stage genome-wide search in psoriasis." Hum Mol Genet **6**(5): 813-820.

Trynka, G., A. Zhernakova, et al. (2009). "Coeliac disease-associated risk variants in TNFAIP3 and REL implicate altered NF-kappaB signalling." Gut **58**(8): 1078-1083.

Turner, J. R. (2009). "Intestinal mucosal barrier function in health and disease." Nat Rev Immunol **9**(11): 799-809.

van Heel, D. A., S. A. Fisher, et al. (2004). "Inflammatory bowel disease susceptibility loci defined by genome scan meta-analysis of 1952 affected relative pairs." Hum Mol Genet **13**(7): 763-770.

Veal, C. D., F. Capon, et al. (2002). "Family-based analysis using a dense single-nucleotide polymorphism-based map defines genetic variation at PSORS1, the major psoriasis-susceptibility locus." Am J Hum Genet **71**(3): 554-564.

Venter, J. C., M. D. Adams, et al. (2001). "The sequence of the human genome." Science **291**(5507): 1304-1351.

Vivier, E., D. H. Raulet, et al. (2011). "Innate or adaptive immunity? The example of natural killer cells." Science **331**(6013): 44-49.

Wallace, C., D. J. Smyth, et al. (2010). "The imprinted DLK1-MEG3 gene region on chromosome 14q32.2 alters susceptibility to type 1 diabetes." Nat Genet **42**(1): 68-71.

Wang, H., T. Peters, et al. (2006). "Activated macrophages are essential in a murine model for T cell-mediated chronic psoriasiform skin inflammation." J Clin Invest **116**(8): 2105-2114.

Wang, Y., S. Sugita, et al. (2000). "The RIM/NIM family of neuronal C2 domain proteins. Interactions with Rab3 and a new class of Src homology 3 domain proteins." J Biol Chem **275**(26): 20033-20044.

Wasserman, N. F., I. Aneas, et al. (2010). "An 8q24 gene desert variant associated with prostate cancer risk confers differential in vivo activity to a MYC enhancer." Genome Res **20**(9): 1191-1197.

Wolters, M. (2005). "Diet and psoriasis: experimental data and clinical evidence." Br J Dermatol **153**(4): 706-714.

Wu, C., C. Orozco, et al. (2009). "BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources." Genome Biol **10**(11): R130.

Wu, M., L. G. Xu, et al. (2003). "SINK is a p65-interacting negative regulator of NF-kappaB-dependent transcription." J Biol Chem **278**(29): 27072-27079.

Yamazaki, K., D. McGovern, et al. (2005). "Single nucleotide polymorphisms in TNFSF15 confer susceptibility to Crohn's disease." Hum Mol Genet **14**(22): 3499-3506.

Young, C., M. H. Allen, et al. (2003). "A Crohn's disease-associated insertion polymorphism (3020insC) in the NOD2 gene is not associated with psoriasis vulgaris, palmo-plantar pustular psoriasis or guttate psoriasis." Exp Dermatol

12(4): 506-509.

Zeggini, E., M. N. Weedon, et al. (2007). "Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes." Science **316**(5829): 1336-1341.

Zhang, X. J., W. Huang, et al. (2009). "Psoriasis genome-wide association study identifies susceptibility variants within LCE gene cluster at 1q21." Nat Genet **41**(2): 205-210.

Psoriasis is associated with pleiotropic susceptibility loci identified in type II diabetes and Crohn disease

N Wolf, M Quaranta, N J Prescott, et al.

J Med Genet 2008 45: 114-116 originally published online November 9, 2007
doi: 10.1136/jmg.2007.053595

Psoriasis is associated with pleiotropic susceptibility loci identified in type II diabetes and Crohn disease

N Wolf,[1] M Quaranta,[1] N J Prescott,[1] M Allen,[1] R Smith,[2,3] A D Burden,[4] J Worthington,[3] C E M Griffiths,[2] C G Mathew,[1] J N Barker,[1] F Capon,[1] R C Trembath[1]

[1] Division of Genetics and Molecular Medicine, King's College London, London, UK; [2] Dermatological Sciences, University of Manchester, Manchester, UK; [3] Arc Epidemiology Unit, University of Manchester, Manchester, UK; [4] Glasgow Western Infirmary, Glasgow, UK

Correspondence to: Professor R C Trembath, Division of Genetics and Molecular Medicine, King's College School of Medicine, 9th Floor Guy's Tower, Guy's Hospital, SE1 9RT London, UK; richard.trembath@genetics.kcl.ac.uk

Received 1 August 2007
Revised 11 October 2007
Accepted 19 October 2007
Published Online First 9 November 2007

ABSTRACT

Background: Psoriasis is an immune-mediated skin disorder that is inherited as a multifactorial trait. Linkage analyses have clearly mapped a primary disease susceptibility locus to the major histocompatibility complex (MHC) region on chromosome 6p21. More recently, whole-genome association studies have identified two non-MHC disease genes (*IL12B* and *IL23R*), both of which also confer susceptibility to Crohn disease (CD).

Objective and methods: To ascertain the genetic overlap between these two inflammatory conditions further, we investigated 15 CD-associated loci in a psoriasis case–control dataset.

Results: The analysis of 1256 patients and 2938 unrelated controls found significant associations for loci mapping to chromosomes 1q24 (rs12035082, $p = 0.009$), 6p22 (rs6908425, $p = 0.00015$) and 21q22 (rs2836754, $p = 0.0003$). Notably, the marker showing the strongest phenotypic effect (rs6908425) maps to *CDKAL1*, a gene also associated with type 2 diabetes.

Conclusions: These results substantiate emerging evidence for a pleiotropic role for s genes that contribute to the pathogenesis of immune-mediated disorders.

Psoriasis is a chronic, immune-mediated skin disorder that affects up to 4% of the population.[1]

Familial recurrence of the disease is well documented, and psoriasis has long been regarded as a multifactorial trait. Linkage studies have repeatedly identified a primary disease susceptibility locus (*PSORS1*), lying within the major histocompatibility complex (MHC) on chromosome 6p21.[2] Several putative non-MHC loci with smaller effects (*PSORS2–10*) have also been described.[3] These regions tend to overlap with genomic intervals conferring susceptibility to other inflammatory conditions, notably Crohn disease (CD).[4 5] The observation of an increased incidence of psoriasis among patients with CD[6] is consistent with these findings and suggests the existence of genetic determinants predisposing to both inflammatory conditions. We and others have previously shown that CD-associated sequence variants in the *IL23R* and *IL12B* genes confer a significant increase in psoriasis risk.[7 8] In this study, we investigated the contribution of CD genetic determinants to psoriasis susceptibility. We assessed 15 CD susceptibility loci recently identified by genome-wide association analysis, and identified three novel and significant disease associations with psoriasis.

METHODS

Subjects

In total, 1256 patients (645 male and 611 female patients) were recruited within the UK through St John's Institute of Dermatology, London (n = 638), Glasgow Western Infirmary (n = 211) and the Dermatology Centre, University of Manchester (n = 407). All patients were of north European descent and had early-onset (occurring before 40 years of age) psoriasis vulgaris. Less than 1% of the patients (n = 8) also had CD. These patients were matched with 2938 controls (1488 male, 1450 female patients), previously analysed by the Wellcome Trust Case–Control Consortium (WTCCC).[9] All patients and controls gave informed consent for the use of their DNA in genetic epidemiology analyses. The study was approved by the Guy's and St Thomas' Hospitals Ethics Committee of Kings College London, the Salford and Trafford Local Research Ethics Committee and North Glasgow University Hospitals NHS Trust Local Research Ethics Committee.

Genotyping and statistical analyses

Patients were typed using TaqMan assays (Applied Biosystems, Foster City, California, USA). Control genotypes were obtained from publicly released data (available at www.wtccc.org.uk/info/summary_stats.shtml). The allele frequencies of cases and controls were compared using a χ^2 test with one degree of freedom, after confirming that the genotype distributions did not deviate from Hardy–Weinberg equilibrium ($p>0.05$). Regression analysis was carried out using PLINK software.[10]

RESULTS

We examined 15 SNPs that had been associated with CD susceptibility in the WTCCC scan[9] and its follow-up study,[11] based on the observation of p values $<10^{-5}$. We did not analyse any single-nucleotide polymorphism (SNP) from the *CARD15/NOD2* locus, as we had previously shown that psoriasis is not associated with sequence variants in this gene.[12] We also excluded chromosome 6p21 markers from our study, as linkage disequilibrium with *PSORS1* alleles would have confounded the interpretation of association signals.

We performed association analyses by comparing the allele frequencies of our patients with those of the WTCCC controls. We applied an assumption that cases could be matched with controls that had been typed on a different platform (the

Table 1 Association analysis results

Marker	Chromosome	Gene	Minor allele frequency		OR	p Value
			Cases	Controls		
rs12035082	1q24	–	0.42 (993/2366)	0.39 (2281/5868)	1.14	0.009
rs10801047	1q31	–	0.07 (164/2338)	0.07 (395/5860)	1.04	0.66
rs10210302	2q37	*ATG16L1*	0.48 (1157/2386)	0.48 (2822/5872)	1.02	0.71
rs9858542	3p21	Many*	0.29 (692/2360)	0.28 (1652/5862)	1.06	0.30
rs17234657	5p13	–	0.13 (319/2448)	0.12 (731/5866)	1.05	0.47
rs6596075	5q23	Many*	0.17 (396/2362)	0.17 (972/5870)	0.98	0.82
rs1000113	5q33	*IRGM*	0.08 (183/2288)	0.07 (394/5854)	1.20	0.045
rs6908425	6p22	*CDKAL1*	0.19 (452/2358)	0.23 (1348/5864)	1.26	0.00015
rs6927210	6q23	–	0.50 (1200/2388)	0.48 (2825/5866)	0.92	0.08
rs887822	7q36	–	0.27 (655/2384)	0.29 (1695/5868)	1.07	0.20
rs6601764	10p15	–	0.42 (973/2316)	0.41 (2368/5810)	1.07	0.20
rs10761659	10q21	–	0.47 (1103/2326)	0.46 (2703/5866)	1.05	0.27
rs10883365	10q24	*NKX2–3*	0.49 (1081/2218)	0.47 (2730/5800)	0.93	0.18
rs2542151	18p11	*PTPN2*	0.17 (372/2140)	0.16 (957/5870)	1.08	0.25
rs2836754	21q22	–	0.39 (847/2168)	0.35 (2072/5862)	1.17	0.0003

*The associations observed for the 3p21 and 5q23 regions spanned large genomic regions, including multiple genes.

Affymetrix 500 K array set), having shown a 99.7% concordance rate between the WTCCC genotypes and those generated by Taqman assays.[11]

Following χ^2 analysis, we observed significant disease associations for three independent markers rs12035082 ($p = 0.009$; OR = 1.14; 95% CI 1.03 to 1.25), rs6908425 ($p = 0.00015$; OR = 1.26; 95% CI 1.12 to 1.42) and rs2836754 ($p = 0.0003$; OR = 1.17; 95% CI 1.06 to 1.30). We also found an association of marginal significance for SNP rs1000113 ($p = 0.045$; OR = 1.20; 95% CI 1.00 to 1.45). In order to address the issue of multiple testing, we calculated false discovery rates (FDR)[13] for the analysis of 15 SNPs. We found that all three significant associations surpassed the 5% FDR threshold ($p \leqslant 0.01$), with two (rs6908425 and rs2836754) exceeding 1% FDR ($p<0.001$).

We next carried out logistic regression analysis, using age and sex as specific covariates. This confirmed that the significance of our association results was not affected by either of these variables.

In order to assess the presence of epistasis between the various disease-associated alleles, we implemented a case-only test for the detection of gene–gene interaction.[14] Under the null hypothesis of no interaction, the genotype frequencies of unlinked markers are expected to be independent from each other. We therefore analysed the distribution of patient genotypes by carrying out χ^2 tests on standard contingency tables. We observed non-significant χ^2 values for all pair-wise comparisons between disease-associated SNPs (rs12035082 *v* rs6908425; rs12035082 *v* rs2836754; rs6908425 *v* rs2836754). We used the same approach to test for interactions with HLA-Cw*0602, the strongest determinant of disease risk from the *PSORS1* locus. The p values for the HLA-C *v* rs6908425 and HLA-C *v* rs2836754 comparisons were non-significant, but the HLA-C *v* rs12035082 analysis generated suggestive evidence for an interaction between the two loci ($p = 0.02$).

DISCUSSION

The existence of shared genetic determinants between psoriasis and CD is clearly documented, with variants in the *IL23R* and *IL12B* genes showing highly significant associations with both diseases.[7 8] With this study we have further investigated the role of CD genes in psoriasis pathogenesis, by examining a panel of CD susceptibility loci, recently identified by whole-genome association analysis. In selecting markers for genotyping, we used a conservative level of statistical significance ($p<10^{-5}$), as several studies have demonstrated the merit of following association signals of this order of magnitude.[11 15 16] Our results further validate this approach, as we were able to document associations with psoriasis for three of the examined SNPs.

The possibility that our findings are due to population stratification is unlikely, given that cases and controls have the same ethnic and geographical origin. Moreover, the regional variation data generated by the WTCCC (available at http://www.b58cgene.sgul.ac.uk/) show that the allele frequencies of SNPs rs12035082, rs6908425 and rs2836754 are homogeneous across the UK, ruling out the possibility that the association is due to hidden population structure.

As our study is based on the same set of controls that allowed the identification of CD susceptibility genes in the WTCCC association scan, we cannot formally exclude the possibility that the similarity of findings may reflect a control sampling effect. However, the fact that the population frequencies of SNP rs12035082 and rs2836754 have been validated in another UK sample[11] argues against this hypothesis.

Given the multifactorial nature of psoriasis inheritance, we implemented a series of case-only tests to assess the presence of epistasis between disease-associated loci. We found suggestive evidence for an interaction between rs12035082 and HLA-C, a finding that should be interpreted with caution, given its marginal significance ($p = 0.02$) and the degree of multiple testing inherent in interaction analyses. Validation of such a modest effect will probably require analysis of a substantially larger patient cohort.

All three disease-associated SNPs map to non-coding regions. Rs6908425, which lies within intron 5 of the *CDKAL1* transcripts, is the only disease-associated SNP found within a known gene. Although the function of *CDKAL1* is unknown, intron 5 variants have been associated with both type 2 diabetes[17] and CD,[9] suggesting that the gene may play an important role in the pathogenesis of both conditions. Given the high frequency of type 2 diabetes, it is probable that a significant percentage of our psoriatic patients (5–10%) may also exhibit diabetes. Although this is undoubtedly a confounding factor, it is unlikely that the results obtained for SNP rs6908425 simply reflect an association with type 2 diabetes. In fact, our calculations indicate that a sample including up to 20% of this patient cohort would have very limited power (~15%)

Key points

- Psoriasis is an inflammatory skin disorder that is inherited as a multifactorial trait. Previous studies have demonstrated that two Crohn disease (CD) susceptibility genes (*IL23R* and *IL12B*) are also associated with psoriasis.
- To further investigate the overlap between psoriasis and CD susceptibility determinants, we examined 15 CD-associated single-nucleotide polymorphisms (SNPs) in a dataset of1256 psoriatic patients and 2938 controls.
- We found that three of the examined SNPs are significantly associated with psoriasis. Of note, the marker showing the strongest phenotypic effect maps to a gene previously associated with type 2 diabetes. These results provide further evidence for the existence of pleiotropic genes predisposing to multiple disorders.

to detect an association with type 2 diabetes, at the observed level of significance. An accurate assessment of rs6908425 effect size will eventually require in-depth clinical characterisation of psoriatic cohorts, including metabolic analysis. The remaining disease-associated markers identified in our study map to intergenic regions. SNP rs2836754 lies between the NP_001001692 and C21orf104 anonymous transcripts on chromosome 21q22. SNP rs12035082 lies in a "gene desert" on chromosome 1q24, within a 77-kb linkage disequilibrium (LD) block. Although the LD block does not include any known gene, it encompasses several genomic segments that are highly conserved among mammals and that may regulate the expression of the downstream *TNFSF18* gene. As *TNFSF18* is a mediator of nuclear factor kappa-b signalling and is a key molecule for the maintenance of immunological self-tolerance,[18] it is interesting to speculate that disease-associated alleles may confer susceptibility to psoriasis and CD by altering *TNFSF18* expression levels.

In conclusion, we have shown that psoriasis is associated with sequence variants predisposing to type 2 diabetes and CD. These observations provide increased support for the existence of critical, yet pleiotropic loci, conferring susceptibility to clinically distinct disorders. Further genetic and functional analyses will now be required to characterise fully the role of disease-associated variants.

Acknowledgements: We wish to express our gratitude to the patients and clinicians who have contributed to this study. We also wish to thank S Fisher for her helpful comments and advice.

Funding: The work was supported by a Medical Research Council PhD studentship (NW), a Stiefel Laboratories PhD studentship (RS), a Psoriasis Association PhD studentship (MO) and grants from the Medical Research Council (grant G0601387; RCT, JNB), The Psoriasis Association (JNB, DB, CEMG) and Arthritis Research Campaign (JW). We acknowledge use of genotype data from the British 1958 Birth Cohort DNA collection, funded by the Medical Research Council (grant G0000934) and The Wellcome Trust (grant 068545/Z/02). None of the above sponsors had any involvement in study design and execution, in the writing of the report, or in the decision to submit the paper for publication.

Competing interests: None.

REFERENCES

1. **Griffiths CE,** Barker JN. Pathogenesis and clinical features of psoriasis. *Lancet* 2007;**370**:263–71.
2. **Capon F,** Munro M, Barker J, Trembath R. Searching for the major histocompatibility complex psoriasis susceptibility gene. *J Invest Dermatol* 2002;**118**:745–51.
3. **Capon F,** Trembath RC, Barker JN. An update on the genetics of psoriasis. *Dermatol Clin* 2004;**22**:339–47, vii.
4. **Lee YA,** Ruschendorf F, Windemuth C, Schmitt-Egenolf M, Stadelmann A, Nurnberg G, Stander M, Wienker TF, Reis A, Traupe H. Genomewide scan in german families reveals evidence for a novel psoriasis-susceptibility locus on chromosome 19p13. *Am J Hum Genet* 2000;**67**:1020–4.
5. **Nair RP,** Henseler T, Jenisch S, Stuart P, Bichakjian CK, Lenk W, Westphal E, Guo SW, Christophers E, Voorhees JJ, Elder JT. Evidence for two psoriasis susceptibility loci (HLA and 17q) and two novel candidate regions (16q and 20p) by genome-wide scan. *Hum Mol Genet* 1997;**6**:1349–56.
6. **Lee FI,** Bellary SV, Francis C. Increased occurrence of psoriasis in patients with Crohn's disease and their relatives. *Am J Gastroenterol* 1990;**85**:962–3.
7. **Capon F,** Di Meglio P, Szaub J, Prescott NJ, Dunster C, Baumber L, Timms K, Gutin A, Abkevic V, Burden AD, Lanchbury J, Barker JN, Trembath RC, Nestle FO. Sequence variants in the genes for the interleukin-23 receptor (IL23R) and its ligand (IL12B) confer protection against psoriasis. *Hum Genet* 2007;**122**:201–6.
8. **Cargill M,** Schrodi SJ, Chang M, Garcia VE, Brandon R, Callis KP, Matsunami N, Ardlie KG, Civello D, Catanese JJ, Leong DU, Panko JM, McAllister LB, Hansen CB, Papenfuss J, Prescott SM, White TJ, Leppert MF, Krueger GG, Begovich AB. A large-scale genetic association study confirms IL12B and leads to the identification of IL23R as psoriasis-risk genes. *Am J Hum Genet* 2007;**80**:273–90.
9. **The Wellcome Trust Case-Control Consortium**. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007;**447**:661–78.
10. **Purcell S,** Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;**81**:559–75.
11. **Parkes M,** Barrett JC, Prescott NJ, Tremelling M, Anderson CA, Fisher SA, Roberts RG, Nimmo ER, Cummings FR, Soars D, Drummond H, Lees CW, Khawaja SA, Bagnall R, Burke DA, Todhunter CE, Ahmad T, Onnie CM, McArdle W, Strachan D, Bethel G, Bryan C, Lewis CM, Deloukas P, Forbes A, Sanderson J, Jewell DP, Satsangi J, Mansfield JC, Cardon L, Mathew CG. Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility. *Nat Genet* 2007;**39**:830–2.
12. **Young C,** MH Allen, A Cuthbert, M Ameen, C Veal, J Leman, AD Burden, B Kirby, CEM Griffiths, RC Trembath, CG Mathew, JNWN Barker. A Crohn's disease-associated insertion polymorphism (3020insC) in the NOD2 gene is not associated woth psoriasis vulgaris, palmo-plantar pustular psoriasis or guttate psoriasis. *Expl Dermatol* 2003;**12**:506–9.
13. **Benjamini Y,** Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Statist Soc B* 1995;**57**:289–300.
14. **Yang Q,** Khoury MJ, Sun F, Flanders WD. Case-only design to measure gene-gene interaction. *Epidemiology* 1999;**10**:167–70.
15. **Rioux JD,** Xavier RJ, Taylor KD, Silverberg MS, Goyette P, Huett A, Green T, Kuballa P, Barmada MM, Datta LW, Shugart YY, Griffiths AM, Targan SR, Ippoliti AF, Bernard EJ, Mei L, Nicolae DL, Regueiro M, Schumm LP, Steinhart AH, Rotter JI, Duerr RH, Cho JH, Daly MJ, Brant SR. Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. *Nat Genet* 2007;**39**:596–604.
16. **Zeggini E,** Weedon MN, Lindgren CM, Frayling TM, Elliott KS, Lango H, Timpson NJ, Perry JR, Rayner NW, Freathy RM, Barrett JC, Shields B, Morris AP, Ellard S, Groves CJ, Harries LW, Marchini JL, Owen KR, Knight B, Cardon LR, Walker M, Hitman GA, Morris AD, Doney AS, McCarthy MI, Hattersley AT. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* 2007;**316**:1336–41.
17. **Zeggini E,** Weedon MN, Lindgren CM, Frayling TM, Elliott KS, Lango H, Timpson NJ, Perry JR, Rayner NW, Freathy RM, Barrett JC, Shields B, Morris AP, Ellard S, Groves CJ, Harries LW, Marchini JL, Owen KR, Knight B, Cardon LR, Walker M, Hitman GA, Morris AD, Doney AS, McCarthy MI, Hattersley AT. Replication of genome-wide association signals in u.k. samples reveals risk loci for type 2 diabetes. *Science* 2007;**316**:1336–41. Epub 2007 Apr 26. Erratum in: Science 2007;**317**:1035–6.
18. **Shimizu J,** Yamazaki S, Takahashi T, Ishida Y, Sakaguchi S. Stimulation of CD25(+)CD4(+) regulatory T cells through GITR breaks immunological self-tolerance. *Nat Immunol* 2002;**3**:135–42.

Genes and Immunity (2009) 10, 654–658

www.nature.com/gene

ORIGINAL ARTICLE

Differential contribution of CDKAL1 variants to psoriasis, Crohn's disease and type II diabetes

M Quaranta[1], AD Burden[2], CEM Griffiths[3], J Worthington[4], JN Barker[1], RC Trembath[1] and F Capon[1]
[1]*King's College London, Division of Genetics and Molecular Medicine, London, UK;* [2]*Department of Dermatology, University of Glasgow, Glasgow, UK;* [3]*Department of Dermatology, University of Manchester, Manchester, UK and* [4]*arc Epidemiology Unit, University of Manchester, Manchester, UK*

Psoriasis is an immune-mediated skin disorder, which is inherited as a complex trait. Genome-wide linkage and association studies have identified a major disease susceptibility locus on chromosome 6p21, as well as a number of genetic determinants of smaller effect. Our group has also documented a significant association between psoriasis and CDKAL1, a gene previously implicated in the pathogenesis of Crohn's disease (CD) and type II diabetes (TIID). With this study, we validate this association, through the analysis of CDKAL1 single nucleotide polymorphism (SNP) rs6908425 in an independently ascertained psoriasis dataset (replication sample: 1323 cases vs 1368 controls, $P = 0.00012$, odds ratio (OR): 1.28; combined sample: 2579 cases vs 4306 controls, $P = 4 \times 10^{-6}$, OR: 1.26). We also show that the association with psoriasis and CD is completely independent from that with TIID. Finally, we report the results of expression studies demonstrating that CDKAL1 transcripts are virtually absent from skin keratinocytes, but are abundantly expressed in immune cells, especially in CD4+ and CD19+ lymphocytes. It is to be noted that our data indicate that CDKAL1 becomes markedly downregulated when immune cells are activated with proliferating signals. Taken together, our results document the presence of allelic heterogeneity at the CDKAL1 locus and suggest that CDKAL1 alleles may confer susceptibility to clinically distinct disorders through differential effects on disease-specific cell types.
Genes and Immunity (2009) **10,** 654–658; doi:10.1038/gene.2009.51; published online 9 July 2009

Keywords: *psoriasis; Crohn's disease; type II diabetes; CDKAL1*

Introduction

Psoriasis is a chronic immune-mediated skin disease, characterized by altered keratinocyte differentiation, angiogenesis and infiltration of inflammatory elements (T-lymphocytes, neutrophils and dendritic cells) into the dermis and the epidermis.[1] Familial recurrence of the disease is well documented, so that psoriasis is widely regarded as complex genetic trait, resulting from the interaction between environmental triggers and inherited disease susceptibility alleles.[2] Genome-wide linkage scans have unambiguously mapped a primary disease susceptibility locus (PSORS1) to the major histocompatibility complex (MHC),[3] where refinement studies and resequencing efforts point to HLA-C as the most likely PSORS1 candidate.[4,5] More recently, the advent of genome-wide association scans has allowed the identification of several non-MHC susceptibility genes, including *IL12B*, *IL23R*, *RNF114*, *TNFAIP3*, *TNIP1*, *IL4/IL13* and *LCE3B/3C*[6–11] Of these, *IL12B* and *IL23R* harbour variants that also confer susceptibility to Crohn's disease (CD)[12] and ankylosing spondylitis.[13] Similarly, *TNFAIP3* alleles have been associated with rheumatoid arthritis,[14] systemic lupus erythematosus[15] and type I diabetes.[16] Similar instances of genes conferring susceptibility to multiple conditions (for example, *PTPN2*, *IL2RA* and *IL2*) have emerged from other genome-wide association scans,[17,18] highlighting unexpected connections between clinically distinct disorders.

In this context, our group has focused on the overlap between the genetic determinants for psoriasis and CD, based on the common inflammatory nature of these conditions, both of which are driven by the infiltration of CD4+ T-lymphocytes in epithelial tissues. Through the analysis of a well-characterized case-control resource, we have previously documented significant associations between psoriasis and three CD susceptibility alleles, mapping to chromosomes 1q24, 6p22 and 21q22.[19] It is to be noted that the marker showing the strongest phenotypic effect is mapped to *CDKAL1*, a gene also associated with type II diabetes (TIID).[20–22] All three loci were recently analysed in a North-American cohort and tentative support was obtained for the *CDKAL1* association ($P = 0.06$; one-sided P-value: 0.03).[23] With this study we provide significant replication data supporting the involvement of *CDKAL1* in the pathogenesis of psoriasis. We also show that this association is independent from

Correspondence: Professor RC Trembath, King's College London, Division of Genetics and Molecular Medicine, 8th floor Tower wing, Guy's Hospital, Great Maze Pond, London SE1 9RT, UK.
E-mail: Richard.Trembath@genetics.kcl.ac.uk
Received 10 April 2009; revised and accepted 5 June 2009; published online 9 July 2009

the allele conferring susceptibility to TIID. Finally we provide real-time PCR data, indicating that *CDKAL1* is strongly expressed in CD4+ T-lymphocytes, a T-cell subset playing a major role in the pathogenesis of both psoriasis and CD.

Materials and methods

Patients

The United Kingdom case-control sample analysed in this study has been described elsewhere.[19] Briefly, a total of 962 British patients of North-European descent were recruited through St John's Institute of Dermatology, London; Glasgow Western Infirmary; and the Dermatology Centre, University of Manchester. All patients were affected by an early onset of psoriasis vulgaris (disease onset occurring before 40 years of age). Control genotypes were obtained by querying the publicly available data generated by the Wellcome Trust Case-Control Consortium (WTCCC).[18] Genotypes from the Collaborative Association Study of Psoriasis (CASP) were obtained from the Genetic Association Information Network (GAIN) Database, through dbGaP accession number phs000019.v1.p1. Primary keratinocytes were obtained from the skin of unaffected individuals, who were undergoing abdominoplasty at Guy's and St Thomas Hospital NHS Trust. All patients and controls gave their informed consent for participating in this project. The study was approved by the Guy's and St Thomas' Hospitals Ethics Committee, the Salford and Trafford Local Research Ethics Committee, and North Glasgow University Hospitals NHS Trust Local Research Ethics Committee.

Genotyping and direct sequencing

Single nucleotide polymorphism (SNP) rs10946398 was typed with a TaqMan SNP genotyping assay (Applied Biosystems, Foster City, CA, USA), run on a 7900HT fast real-time PCR system (Applied Biosystems). We expect the error rate for this assay to be less than 1%, as we have previously shown that our TaqMan genotypes are 99.7% concordant with those obtained by the WTCCC for the same markers.[19]

CDKAL1 RT-PCR products were sequenced with a BigDye Terminator cycle sequencing kit (Applied Biosystems) and loaded on an ABI 3730*xl* automated sequencer (Applied Biosystems). The sequence of the amplified products was compared with that of *CDKAL1* full-length cDNA, using Sequencher (Gene Codes Corporation, Ann Arbor, MI, USA).

Statistical analyses

Genotype quality controls (analysis of call rate and Hardy–Weinberg equilibrium) and χ^2-tests were implemented with PLINK. The genotypes of marker rs6908425 were also imputed with PLINK,[24] using *CDKAL1* HapMap data as a reference and applying the following QC filters: call rate >90%, MAF$\geqslant$1% and HWE *P*-value$\geqslant 10^{-4}$. The value of the INFO metric, which varies from 0 to 1 and measures the quality of imputed genotypes, was 0.92. The results of this analysis were validated with the MACH software,[25] which confirmed the presence of significant association between psoriasis and imputed rs6908425 genotypes (data not shown). Meta-analyses of the United Kingdom and CASP samples were implemented with Review Manager. The statistical power of the examined datasets was assessed with Genetic Power Calculator,[26] using published allele frequencies and odds ratios.

Expression studies

Multiple tissues and human blood fractions cDNA panels were purchased from Clontech (Mountain View, CA, USA). Both panels contain first-strand cDNA preparations obtained from pools of tissues/cells. Data provided by the manufacturer and based on the immuno-staining of relevant surface markers (CD4, CD8, CD19, CD25 and CD171), indicate that the purity of the lymphocyte populations in the blood fraction panel exceeds 95% for resting cells and 84% for activated cells.

Primary keratinocytes were grown in Epilife medium (Invitrogen, Portland, OR, USA) and were treated with 100 ng ml^{-1} of each cytokine (all purchased from R&D Systems, Minneapolis, MN, USA) at passages 2–4. The cytokines bioactivity was verified by confirming their ability to induce the expression of known target genes (data not shown). About 24 h after cytokine treatment (5 h for interferon (IFN)-α), cells were harvested, total RNA was extracted with TRI reagent (Sigma-Aldrich, St Louis, MO, USA) and cDNA was reverse transcribed using a Superscript first-strand synthesis kit (Invitrogen).

Transcripts were quantified by using TaqMan gene expression assays (Applied Biosystems), according to the manufacturer's protocol. Briefly, 2 µl of cDNA were amplified in a final volume of 20 µl, in the presence of primers and probes for *CDKAL1* (Applied Biosystems assay id: Hs00214949_m1), *PPIA* (house-keeping gene encoding cyclophilin A; assay id: Hs99999904_m1) or *RPLPO* (house-keeping gene encoding large PO ribosomal protein; assay id: Hs99999902_m1). All real-time PCR reactions were carried out in duplicate, using the following cycling conditions: 15 min 95°C; (30 s 95°C, 1 in 60°C) $\times$ 40. After amplification, relative *CDKAL1* expression levels were derived using the $\Delta\Delta$Ct method.[27]

Results

Validation of CDKAL1 *association with psoriasis*

In the first phase of this study, we sought access to the data generated by the CASP (an association scan of 1409 patients and 1436 controls of European ancestry[10]), with a view to validating the association between psoriasis and the CD loci on chromosomes 1q24, 6p22 and 21q22. Good quality SNP genotypes (call rate >90%, *P*-value for Hardy–Weinberg equilibrium >0.05) were available for markers rs12035082 (chromosome 1q24) and rs2836754 (chromosome 21q22), in 1323 cases and 1368 controls. As SNP rs6908425 (chromosome 6p22) was not present in the Perlegen array used by CASP, genotypes for this marker were imputed using PLINK. After association analysis, the CASP results were compared with those of our original study. As shown in Table 1, *CDKAL1* SNP rs6908425 was the only marker generating convincing association in both the discovery and the replication sample. No evidence of heterogeneity was detected between the United Kingdom and CASP datasets ($P = 0.85$; $I^2 = 0\%$) and a meta-analysis of the two studies yielded a *P*-value of 4×10^{-6} (OR: 1.26; 95% CI:

Table 1 Association analysis of 1q24, 6p22 and 21q22 CD loci in two psoriasis datasets

Marker (chromosome)	*Gene*	*Discovery sample (1256 cases vs. 2938 controls)*			*CASP sample (1323 cases vs 1368 controls)*			*Combined sample (2579 cases vs 4306 controls)*		
		Case MAF	*Control MAF*	P-value	*Case MAF*	*Control MAF*	P-value	P-value	*Odds ratios (95% CI)*	*Risk allele*
rs12035082 (1q24)	—	0.42	0.39	0.009	0.40	0.41	0.58	0.05	1.07 (1.00–1.15)	—
rs6908425 (6p22)	*CDKAL1*	0.19	0.23	0.00015	0.20	0.24	0.00012	**4×10^{-6}**	**1.26 (1.15–1.41)**	C
rs2836754 (21q22)	*FLJ45139*	0.39	0.35	0.0003	0.39	0.36	0.06	0.001	1.15 (1.06–1.25)	—

Abbreviations: CASP, Collaborative Association Study of Psoriasis; CD, Crohn's disease; CI, confidence interval; MAF, minor allele frequency. The details relating to the *CDKAL1* association are highlighted in bold.

1.15–1.41). A genotypic analysis of the combined resource (2579 cases vs 4306 controls) further confirmed the association between rs6908425 and psoriasis (*P*-value for Armitage trend test: 1.4×10^{-8}).

The CDKAL1 SNP associated with TIID does not confer susceptibility to psoriasis or CD

The data generated by the HapMap Consortium indicate that rs6908425 is not correlated with the *CDKAL1* markers that have been associated with TIID (rs10946398, rs9465871, rs7767391; $r^2 \leqslant 0.04$ for all), suggesting that the two association signals are completely independent. This does not exclude the possibility that TIID associated SNPs may confer susceptibility to psoriasis, independently from rs6908425. To explore this hypothesis, we analysed the SNP yielding the strongest association with TIID (rs10946398), in 962 United Kingdom psoriatic cases and 2938 unrelated controls. We also examined the data generated by the CASP study for SNP rs9460546 (r^2 with rs10946398: 1.0). Although our calculations indicated that the statistical power of the combined dataset exceeded 90%, we failed to detect any significant association between rs10943698 and psoriasis ($P = 0.60$). We next queried the data generated by the Wellcome Trust Case Control Consortium,[18] to assess whether rs10943698 was associated with CD. Once more, the analysis of a sizeable resource (1747 cases and 2938 controls) failed to generate any evidence for association ($P = 0.84$). Taken together, these data document the presence of allelic heterogeneity at the *CDKAL1* locus, where rs6908425 shows an association with psoriasis and CD that is completely independent from that between rs10946398 and TIID.

CDKAL1 *transcripts are virtually absent from skin keratinocytes, but are clearly expressed in immune cells*

The function of the CDKAL1 protein is currently unknown and only limited information is available on *CDKAL1* gene expression, as previous studies have mostly focused on tissues, which are relevant to the pathogenesis of diabetes.[20,22] We therefore carried out a comprehensive analysis of *CDKAL1* transcript levels, by real-time PCR analysis of cDNAs from 20 tissues and cell types. In agreement with published data, we observed strong *CDKAL1* expression in skeletal muscle and weaker signals in the pancreas, the liver and the kidney (Figure 1). Surprisingly, very low transcript levels were observed in some of the tissues affected by CD (the small intestine and the colon) and psoriasis (skin keratinocytes). Conversely, *CDKAL1* was abundantly expressed in immune cells, with the strongest signal seen in CD4+ T-lymphocytes and CD19+ B-lymphocytes (Figure 1).

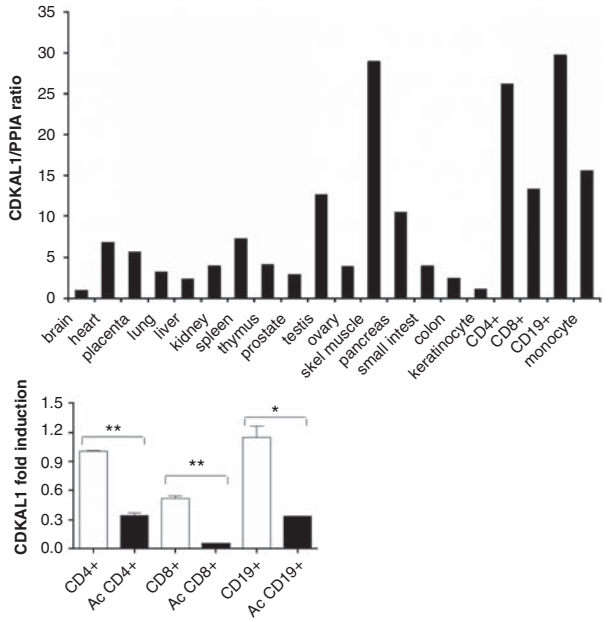


Figure 1 Real-time PCR analysis of *CDKAL1* gene expression. Top panel: Transcript levels were quantified in 20 tissues and cell types, using the expression of *PPIA* as an internal control. The *CDKAL1*/*PPIA* ratio for the brain was set as a baseline value, to which all transcript levels were normalized. Skel. Muscle: skeletal muscle; small intest: small intestine; CD4+: CD4+ T-lymphocytes; CD8+: CD8+ T-lymphocytes; CD19+: CD19+ B-lymphocytes. Bottom panel: Transcript levels were quantified in cDNAs obtained from pools of resting and activated lymphocytes, using the expression of *PPIA* as an internal control. The *CDKAL1*/*PPIA* ratio for CD4+ T-lymphocytes was set as a baseline value, to which all transcript levels were normalized. Error bars refer to technical duplicates in real-time PCR experiments. CD4+: resting CD4+ T-lymphocytes; Ac CD4+: activated CD4+ T-lymphocytes; CD8+: resting CD8+ T-lymphocytes; Ac CD8+: activated CD8+ T-lymphocytes; CD19+: resting CD19+ B-lymphocytes; Ac CD19+: activated CD19+ B-lymphocytes. *$P < 0.05$; **$P < 0.01$.

The information available on the Ensembl genome browser indicates the existence of four *CDKAL1* transcript isoforms (the full length cDNA, two transcripts lacking portions of the gene untranslated regions and a shorter cDNA lacking exons 4 and 13), which can all be detected by the exon 10 probe used for our real-time PCR

experiments. To determine whether our expression analysis had detected the full-length *CDKAL1* transcript, the isoform lacking exons 4 and 13, or both, we designed RT-PCR primers that would amplify both transcripts, by annealing to exons 1 and 16. After electrophoresis of RT-PCR products, we consistently observed a single amplification band, whose molecular weight and sequence matched those of the full-length *CDKAL1* transcript (Supplementary Figure 1). This led us to conclude that the short *CDKAL1* isoform was not expressed in any of the examined tissues, so that the real-time PCR signals we had previously observed were because of the presence of transcripts that included all of the *CDKAL1* coding exons.

Given the fundamental role played by lymphocytes and keratinocytes in the pathogenesis of psoriasis, we decided to further investigate *CDKAL1* expression in these cell types. We first analysed cDNA from primary keratinocytes obtained from healthy donors and treated with a panel of pro-inflammatory cytokines (IFNα, IFNγ, TNFα, IL-17, IL-22 and IL-23), to assess whether any of these molecules could induce *CDKAL1* gene expression. We failed to observe up-regulation of gene expression by any of these cytokines (Supplementary Figure 2).

We next focused our attention on lymphocytes and carried out real-time PCR analyses of T- and B-cells that had been activated, with well-characterized mitogens (concavalin A for CD4+ lymphocytes, phytohaemoagglutinin for CD8$^+$ lymphocytes and pokeweed mitogen for CD19+ lymphocytes, all used for 72 h). The comparison with resting cells showed that *CDKAL1* is markedly downregulated in CD4+, CD8+ and CD19+ proliferating lymphocytes (Figure 1). To exclude the possibility that these data might reflect a differential expression of the housekeeping gene (*PPIA*) in stimulated vs non-stimulated cells, we also measured the transcription of *RPLPO* (ribosomal protein, large PO) mRNA. The amplification of this additional reference gene confirmed our previous results, indicating that *CDKAL1* is genuinely downregulated in proliferating lymphocytes (Supplementary Figure 3).

Discussion

The identification of shared genetic determinants for clinically distinct disorders is an emerging theme underlying the results of recent genome-wide association scans.[28] With this study, we sought to dissect the molecular genetic basis of one such observation of overlap between disease susceptibility loci. *CDKAL1* was identified as a putative susceptibility gene for TIID and CD by the Wellcome Trust Case Control Consortium.[18] Both associations were extensively replicated by follow-up studies,[20,22,29] independent scans[21] and analyses of related phenotypes (for example, pancreatic β-cell function for TIID[30] and ulcerative colitis for CD[31]). Our group subsequently identified a significant association between *CDKAL1* and psoriasis, in a sizeable cohort of British descent.[19] Here, we provide an independent validation for *CDKAL1* involvement in psoriasis, through the analysis of data generated by the CASP study. *CDKAL1* is, thus, the third CD gene (after *IL23R* and *IL12B*) to be convincingly implicated in psoriasis susceptibility. This suggests a pathogenetic mechanism, whereby risk alleles at shared disease loci may account for the increased incidence of psoriasis among CD patients.[32,33]

Our results also show that the *CDKAL1* SNP conferring susceptibility to TIID is not associated with psoriasis or CD, thus validating the presence of two independent association signals at this locus. It is to be noted that the psoriasis associated SNP (rs6908425), the TIID associated SNP (rs10946398) and the various markers that are in LD with them (rs4712528, rs7748720 for rs6908425; rs9460546, rs9565871 for rs10946398; $r^2 > 0.8$ for all) all map to *CDKAL1* intron 5, suggesting that this gene region may contain important regulatory elements. This notion is consistent with the annotation data available through the University of California at Santa Cruz (UCSC) genome browser, which shows the presence of evolutionary conserved regions in proximity of both rs6908425 and rs10946398. In this context, it is tempting to speculate that the two SNPs may confer susceptibility to clinically distinct disorders, by affecting regulatory elements that drive *CDKAL1* expression in different cell types. It should be noted that the existence of *CDKAL1* tissue-specific enhancers has already been demonstrated for a conserved segment of intron 9, which can drive the expression of a *lacZ* reporter in the mouse foetal brain (data generated by Lawrence Berkley National Laboratory and available through VISTA whole genome enhancer browser[34]).

As the function of the CDKAL1 protein is unknown, we analysed the gene expression pattern in a wide range of tissues and cell types. Although our real-time PCR analysis confirmed previous reports of high *CDKAL1* expression in the skeletal muscle, it also generated a number of novel and significant findings. *CDKAL1* transcripts were virtually absent from primary keratinocytes, the small intestine and the colon. Conversely, gene expression was clearly detected in immune cells. A very strong signal was observed in CD4$^+$ T-lymphocytes, which are crucial to the pathogenesis of both psoriasis and CD. *CDKAL1* was also found to be markedly downregulated in activated T- and B-cells, providing additional support to the notion that its gene product has an important role in the biology of lymphocytes. Further studies will now be required to dissect the molecular interactions underpinning *CDKAL1* function in the immune system.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

We thank Massimo Mangino for his contribution to the statistical analyses and John Mee for providing access to primary keratinocytes. Funding support for the CASP study was provided by the National Institutes of Health and the genotyping of samples was implemented through the Genetic Association Information Network (GAIN). Samples and associated phenotype data for the CASP study were provided by Goncalo Abecasis. We acknowledge the use of genotype data from the British 1958 Birth Cohort DNA collection, funded by the

Medical Research Council (grant G0000934) and The Wellcome Trust (grant 068545/Z/02). This work was also supported by a PhD studentship from The Psoriasis Association (MQ) and grants from the British Skin Foundation (FC) and the Medical Research Council (JNB, RCT). *Funding*: British Skin Foundation, The Psoriasis Association, Medical Research Council

References

1 Griffiths CE, Barker JN. Pathogenesis and clinical features of psoriasis. *Lancet* 2007; **370**: 263–271.
2 Lowes MA, Bowcock AM, Krueger JG. Pathogenesis and therapy of psoriasis. *Nature* 2007; **445**: 866–873.
3 Capon F, Munro M, Barker J, Trembath R. Searching for the major histocompatibility complex psoriasis susceptibility gene. *J Invest Dermatol* 2002; **118**: 745–751.
4 Nair RP, Stuart PE, Nistor I, Hiremagalore R, Chia NV, Jenisch S *et al.* Sequence and haplotype analysis supports HLA-C as the psoriasis susceptibility 1 gene. *Am J Hum Genet* 2006; **78**: 827–851.
5 Veal CD, Capon F, Allen MH, Heath EK, Evans JC, Jones A *et al.* Family-based analysis using a dense single-nucleotide polymorphism-based map defines genetic variation at PSORS1, the major psoriasis- susceptibility locus. *Am J Hum Genet* 2002; **71**: 554–564.
6 Capon F, Bijlmakers MJ, Wolf N, Quaranta M, Huffmeier U, Allen M *et al.* Identification of ZNF313/RNF114 as a novel psoriasis susceptibility gene. *Hum Mol Genet* 2008; **17**: 1938–1945.
7 Capon F, Di Meglio P, Szaub J, Prescott NJ, Dunster C, Baumber L *et al.* Sequence variants in the genes for the interleukin-23 receptor (IL23R) and its ligand (IL12B) confer protection against psoriasis. *Hum Genet* 2007; **122**: 201–206.
8 Cargill M, Schrodi SJ, Chang M, Garcia VE, Brandon R, Callis KP *et al.* A large-scale genetic association study confirms IL12B and leads to the identification of IL23R as psoriasis-risk genes. *Am J Hum Genet* 2007; **80**: 273–290.
9 de Cid R, Riveira-Munoz E, Zeeuwen PL, Robarge J, Liao W, Dannhauser EN *et al.* Deletion of the late cornified envelope LCE3B and LCE3C genes as a susceptibility factor for psoriasis. *Nat Genet* 2009; **41**: 211–215.
10 Nair RP, Duffin KC, Helms C, Ding J, Stuart PE, Goldgar D *et al.* Genome-wide scan reveals association of psoriasis with IL-23 and NF-kappaB pathways. *Nat Genet* 2009; **41**: 199–204.
11 Zhang XJ, Huang W, Yang S, Sun LD, Zhang FY, Zhu QX *et al.* Psoriasis genome-wide association study identifies susceptibility variants within LCE gene cluster at 1q21. *Nat Genet* 2009; **41**: 205–210.
12 Duerr RH, Taylor KD, Brant SR, Rioux JD, Silverberg MS, Daly MJ *et al.* A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. *Science* 2006; **314**: 1461–1463.
13 Burton PR, Clayton DG, Cardon LR, Craddock N, Deloukas P, Duncanson A *et al.* Association scan of 14 500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. *Nat Genet* 2007; **39**: 1329–1337.
14 Plenge RM, Cotsapas C, Davies L, Price AL, de Bakker PI, Maller J *et al.* Two independent alleles at 6q23 associated with risk of rheumatoid arthritis. *Nat Genet* 2007; **39**: 1477–1482.
15 Graham RR, Cotsapas C, Davies L, Hackett R, Lessard CJ, Leon JM *et al.* Genetic variants near TNFAIP3 on 6q23 are associated with systemic lupus erythematosus. *Nat Genet* 2008; **40**: 1059–1061.
16 Fung EY, Smyth DJ, Howson JM, Cooper JD, Walker NM, Stevens H *et al.* Analysis of 17 autoimmune disease-associated variants in type 1 diabetes identifies 6q23/TNFAIP3 as a susceptibility locus. *Genes Immun* 2009; **10**: 188–191.
17 Hunt KA, Zhernakova A, Turner G, Heap GA, Franke L, Bruinenberg M *et al.* Newly identified genetic risk variants for celiac disease related to the immune response. *Nat Genet* 2008; **40**: 395–402.
18 The Wellcome Trust Case-Control Consortium. Genome-wide association study of 14 000 cases of seven common diseases and 3000 shared controls. *Nature* 2007; **447**: 661–678.
19 Wolf N, Quaranta M, Prescott NJ, Allen M, Smith R, Burden AD *et al.* Psoriasis is associated with pleiotropic susceptibility loci identified in type II diabetes and Crohn disease. *J Med Genet* 2008; **45**: 114–116.
20 Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL *et al.* A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* 2007; **316**: 1341–1345.
21 Steinthorsdottir V, Thorleifsson G, Reynisdottir I, Benediktsson R, Jonsdottir T, Walters GB *et al.* A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. *Nat Genet* 2007; **39**: 770–775.
22 Zeggini E, Weedon MN, Lindgren CM, Frayling TM, Elliott KS, Lango H *et al.* Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* 2007; **316**: 1336–1341.
23 Li Y, Liao W, Chang M, Schrodi SJ, Bui N, Catanese JJ *et al.* Further genetic evidence for three psoriasis-risk genes: ADAM33, CDKAL1, and PTPN22. *J Invest Dermatol* 2009; **129**: 629–634.
24 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D *et al.* PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; **81**: 559–575.
25 Li Y, Ding CJ, Abecasis G. Mach 1.0: Rapid haplotype reconstruction and missing genotype inference. *Am J Hum Genet* 2006; **79**: S2290.
26 Purcell S, Cherny SS, Sham PC. Genetic power calculator: Design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* 2003; **19**: 149–150.
27 Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods (San Diego, California)* 2001; **25**: 402–408.
28 Seldin MF, Amos CI. Shared susceptibility variations in autoimmune diseases: A brief perspective on common issues. *Genes Immun* 2009; **10**: 1–4.
29 Barrett JC, Hansoul S, Nicolae DL, Cho JH, Duerr RH, Rioux JD *et al.* Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat Genet* 2008; **40**: 955–962.
30 Pascoe L, Tura A, Patel SK, Ibrahim IM, Ferrannini E, Zeggini E *et al.* Common variants of the novel type 2 diabetes genes CDKAL1 and HHEX/IDE are associated with decreased pancreatic beta-cell function. *Diabetes* 2007; **56**: 3101–3104.
31 Anderson CA, Massey DC, Barrett JC, Prescott NJ, Tremelling M, Fisher SA *et al.* Investigation of Crohn's disease risk loci in ulcerative colitis further defines their molecular relationship. *Gastroenterology* 2009; **136**: 523–529 e3.
32 Lee FI, Bellary SV, Francis C. Increased occurrence of psoriasis in patients with Crohn's disease and their relatives. *Am J Gastroenterol* 1990; **85**: 962–963.
33 Yates VM, Watkinson G, Kelman A. Further evidence for an association between psoriasis, Crohn's disease and ulcerative colitis. *Br J Dermatol* 1982; **106**: 323–330.
34 Visel A, Minovitsky S, Dubchak I, Pennacchio LA. VISTA Enhancer Browser–a database of tissue-specific human enhancers. *Nucleic Acids Res* 2007; **35**: D88–D92.

Supplementary Information accompanies the paper on Genes and Immunity website (http://www.nature.com/gene)